

Georgia II STEM Higher Education Evaluation Design Report (Revised)

Charles A. Goldman, Rita Karam, Katya Migacheva, Matthew Cefalu, Thomas Goughnour

RAND Education and Labor

PR-4197-MCC

July 2019

Prepared for Millennium Challenge Corporation

RAND's publications do not necessarily reflect the opinions of its research clients and sponsors. **RAND®** is a registered trademark.



Preface

The U.S. Millennium Challenge Corporation (MCC) has made a major investment to bring international higher education to Georgians, particularly in the area of Science, Technology, Engineering, and Math (STEM). Through this investment, San Diego State University has partnered with three Georgian public universities to offer bachelor's degrees in STEM disciplines. The project also aims to increase the capacity of the Georgian public universities to offer internationally-accredited programs in disciplines outside of the spectrum covered by the SDSU-Georgia Partnership.

MCC selected RAND as the independent evaluator for this project. In this report, the RAND team delineates the evaluation design and approach that will be used for evaluating the successes and challenges of this effort. Using a mix of qualitative and quantitative methodologies, RAND will examine such areas as the evolution of the Partnership and the SDSU-Georgia programming specifically, the roles and experiences of different stakeholders, and the impact of the program on the participating students.

This research is being conducted with funding from MCC by RAND Education and Labor, a division of the RAND Corporation that conducts rigorous, objective research to help decisionmakers and practitioners find solutions to education and labor market challenges. For more about RAND Education and Labor, visit <https://www.rand.org/education-and-labor.html>.

Table of Contents

Preface.....	ii
Figures.....	iv
Tables.....	iv
Acknowledgments.....	iv
Abbreviations.....	v
1. Introduction and Background	1
2. Overview of The Compact and The Intervention Evaluated	3
A. Overview of the Project and Implementation Plan	3
B. Logic Model.....	6
C. Initial Cost-Benefit & Beneficiary Analysis.....	9
D. Review of Relevant Documentation.....	10
3. Evaluation Design.....	12
A. Evaluation Questions	12
B. Evaluation Design Overview	14
C. Quantitative Approach.....	16
Design.....	16
Treatment Group	17
Comparison Group	17
Statistical Power Calculations	18
Primary Data Collection.....	19
Piloting the Survey and Entering Data	20
Analysis Plan.....	21
Challenges and Limitations	23
Estimating Rate of Return	23
D. Qualitative approach.....	25
Primary Data Collection.....	28
Analysis Plan.....	29
4. Data Collection Summary and Timeline.....	30
5. Administrative Information	32
Human Subjects Protection	32
Preparing Data Files for Access, Privacy and Documentation.....	32
Dissemination.....	32
Evaluation Team Roles and Responsibilities	32
Evaluation Timeline and Reporting Schedule	33
A. Potential Programs for the Selection of the Comparison Group.....	35
B. Stakeholder Comments and Responses	41
References.....	54

Figures

Figure 2.1. Project Logic Model	8
---------------------------------------	---

Tables

Table 3.1. Evaluation Questions, Key Outcomes, and Data Sources	15
Table 3.2. Power to detect a difference in monthly wages between 800 (comparison) and 1000 (treatment)	19
Table 5.1. Timing of Major Evaluation Activities and Reports	33

Acknowledgments

We are grateful for the support and insights of key staff at MCC and MCA-Georgia in helping us launch this evaluation and frame the research design, in particular Jenny Heintz, Ryan Moore, and Zura Simonia.

We also appreciate the careful reviews of this report by multiple staff from MCC, MCA-Georgia, SDSU and our RAND colleagues Fatih Unlu and Shanthi Nataraj.

Abbreviations

ABET	Accreditation Board for Engineering and Technology
ACS	American Chemical Society
GTU	Georgian Technical University
ICC	intraclass correlation coefficient
IRB	institutional review board
ISU	Ilya State University
MCA-G	Millennium Challenge Account-Georgia
MCC	Millennium Challenge Corporation
SDSU	San Diego State University
STEM	Science, technology, engineering, and mathematics
TSU	Tbilisi State University
WASC	Western Association of Schools and Colleges

1. Introduction and Background

The Georgian government and society realize that progressive higher education responding to the needs of the Georgian labor market is essential to propel the nation's economy toward growth and rapid development. Educating and sustaining highly skilled workers, as well as becoming a destination country for those interested in pursuing high quality educational experiences abroad are prominent goals of Georgia's Ministry of Education and Science.¹

While Georgia's higher education system has undergone significant modernization reforms, aiming at better alignment with and integration into European and Western education networks and at preparation of a highly skilled and innovative workforce, important gaps remain. Georgia lacks sufficient science, technology, engineering, and math (STEM) programs to sustain strategic STEM fields of study and research and to fulfill the demands of the labor market. The Soviet legacy of highly centralized control over the education system has prevented greater synchronization between the Georgian market demands and higher education offerings. In addition, when it comes to STEM education, significant gender gaps still exist: female students represent only about one third of all students enrolled in STEM-related fields of higher education. Finally, if the opportunity is available to them, many young Georgians prefer to get education abroad and those with the best skills tend to seek employment abroad as well.

To address these challenges, the Millennium Challenge Corporation (MCC) has made a major investment aimed at facilitating high quality inclusive university-level STEM education in Georgia. Through this investment, San Diego State University (SDSU) partnered with three Georgian public universities to offer Bachelor's degrees in a range of STEM disciplines to Georgian students. The project has also aimed to increase capacity of the Georgian public universities to offer internationally-accredited programs.

In what follows, we will refer to the agreements, arrangements, contractual ties, and collaboration between the San Diego State University and the three participating Georgian public universities as "the Partnership"; we will refer to all of the activities that developed from MCC's funding for university-level STEM education in Georgia as "the Project." The term Project, thus, includes the joint SDSU-Georgia program as well as all other activities designed to build the infrastructure, capacity and sustainability of partner universities. When discussing the joint degree program between SDSU and Georgia only, we will refer to it in the text as the SDSU-Georgia program or the program.

In this report, the RAND team outlines the proposed design of the evaluation of the Project. Following this introduction, Chapter 2 offers a brief overview of the Project under assessment, its background and logic. Chapter 3 presents the research questions and the proposed methodologies to help answer them. Chapter 4 presents a summary of the data collection activities and the timeline for their execution. Chapter 5 discusses human subjects and data protection, staffing, overall project timelines as well as plans for disseminating the evaluation

¹ Observations in this introduction are drawn from the evaluation solicitation (*Solicitation*), the monitoring and evaluation plan (*Monitoring and Evaluation Plan, Compact II*, Millennium Challenge Account – Georgia, Version 3, January 2018), and the project team's initial interviews (MCC program officer and former and current project leads, MCA Georgia leads).

findings. Appendix A provides details of the survey comparison group and Appendix B provides a record of the stakeholder comments on earlier drafts of this report, and our responses.

2. Overview of The Compact and The Intervention Evaluated

A. Overview of the Project and Implementation Plan²

On July 26, 2013, the United States of America, acting through the Millennium Challenge Corporation (MCC) – an independent U.S. foreign aid agency with a mission to help lead the fight against global poverty – signed a five-year, \$140-million compact with the Government of Georgia, which was further ratified by the Parliament of Georgia. The largest investment in Georgia’s education sector to date, the compact has aimed to develop Georgia’s human capital, advance economic growth and reduce poverty in Georgia. Georgia II STEM Higher Education Project is a part of this compact and involves a 30 million investment aimed at improving the quality of Georgian university education in STEM fields, and thereby fostering a skilled Georgian labor force and increasing Georgians’ earning potential. The Georgia II STEM Higher Education Project is the focus of the present evaluation efforts.

The compact entered into force on July 1, 2014 and will end on June 30, 2019. The Millennium Challenge Account-Georgia (MCA-Georgia, or MCA-G) has been designated by the Georgian government to administer and oversee all of the Georgia II STEM Higher Education Project activities.

The key goal of the Project is the long-term delivery of high-quality STEM Bachelor’s degrees in Georgia. The Project has aimed to attain this goal by:

- bringing a U.S. university to Georgia to partner with Georgian public universities to offer U.S. bachelor’s degree programs in the STEM academic disciplines,
- providing capacity enhancement for Georgian public universities with the goal of Georgian university programs reaching international standards and acquiring international program accreditation,
- improving infrastructure and technical conditions and increasing capacity of Georgian Higher Education Institutions, and
- preparing world-class skilled STEM specialists from diverse backgrounds.

Importantly, because inequality of female and lower-income families’ participation in the labor force can be a significant constraint to economic growth and poverty reduction,³

² This section draws heavily from *Monitoring and Evaluation Plan, Compact II*. Millennium Challenge Account – Georgia, Version 3, January 2018, MCC and MCA-Georgia websites (http://mcageorgia.ge/index.php/main/read_project/3#) and supplemented by observations from initial interviews conducted by the project team.

³ Peter Gay, Nino Javakhishvili, and Giorgi Shubitide, *Studies of Labor Demand, Barriers to Participation in STEM Education Programs and Occupations in Georgia*, IPM Research per request by MCA-Georgia. Tbilisi, 2014. Retrieved from http://mcageorgia.ge/cms/kcfinder/upload/files/Final_Report_2014%2016%2001%20modified%2027Feb2014%281%29.pdf

promoting social and gender integration within all of the Project activities has been an important and overarching objective.⁴ The Project has set out to do so by:

- implementing activities based on the findings and recommendations of studies that identify barriers to female and socially vulnerable students' participation in STEM programs;
- ensuring that higher education programs include specific activities for outreach, mentoring, and career counseling programs directed toward women, low-income students and other disadvantaged populations; and
- offering needs-based scholarships.⁵

In addition, while the Project overall did not have the recruitment of foreign students among its primary objectives, attraction of international students into Georgian university programs has been one of the priorities of the Government of Georgia. MCC and MCA-G worked with SDSU to make its Georgia degree programs accessible to foreign students.

The Partnership

To pursue these objectives, MCC and MCA-Georgia have facilitated the establishment of the Partnership between the San Diego State University and three public Georgian universities (Tbilisi State University, Georgian Technical University, and Ilia State University).

MCA-Georgia signed a 15-month pre-enrollment agreement with San Diego State University (SDSU), under which SDSU undertook the necessary actions to enroll students starting July 2014, followed by a 45-month collaborative agreement to complete the remainder of the project through July 2019. SDSU has been responsible for design, development, and delivery of the academic programs, as well as for the required infrastructure improvements and establishing connections with potential employers. According to this agreement, SDSU was expected to:

- administer and offer academic programs that are professionally (ABET, ACS) and regionally (WASC) accredited and internationally recognized,
- assist partner universities to become ready to apply for accreditation,
- develop curricula and train Georgian faculty,
- engage in outreach to diverse high school students for recruitment,
- develop facilities that deliver the SDSU programs, and
- develop partnerships with industry.

To be able to offer SDSU programming within the existing institutional framework in Georgia, as well as to build local capacity, SDSU has partnered with three public universities in Tbilisi, Georgia. Currently, SDSU is in the process of implementing U.S. and Georgian Bachelor's dual degree programs in six disciplines: chemistry, computer science, computer engineering, electrical engineering, civil engineering, and construction engineering. Students

⁴ Social and gender integration plan, MCA-Georgia, 2014. Retrieved from <http://mcageorgia.ge/cms/kcfinder/upload/files/Social%20and%20Gender%20Integration%20Plan%20Aug%2027%2C%202014%281%29.pdf>

⁵ <http://mcageorgia.ge/index.php/main/gender>

who will graduate from these academic programs will receive two diplomas, from both San Diego State University and one of the partner universities. The partner universities and the SDSU degree programs offered on their premises are listed in the table below.

Table 1. Georgian partner universities and the SDSU degree programs they host

Partner University	SDSU Degree Program
Tbilisi State University	Chemistry Biochemistry Computer Engineering Electrical Engineering Computer Science
Ilia State University	Computer Engineering Electrical Engineering
Georgia Technical University	Computer Engineering Electrical Engineering Civil Engineering Construction Engineering

Participants

Students are the key targeted participants of the Project. The Project has aimed to recruit highly qualified high school Georgian students interested in STEM, some of whom might have pursued education abroad, had they not had the opportunity obtaining a SDSU degree. This group of students were targeted for improving graduates' STEM skills to match employer needs in Georgia, increase graduates' income and provide them with opportunities to pursue graduate education. The Project has implemented efforts to target women into the program and international students.

Other stakeholder groups have been involved in the design and delivery of the Project or provision of support, but are not considered participants. The RAND evaluation design will collect information from the following stakeholder groups to understand the nature of the Partnership and programming (SDSU-Georgia program) provided for the purpose of informing participants' outcomes (successes or failures).

- MCC: Provide funding to establishing the Project, technical assistance, and assistance with gaining cooperation of key stakeholders in Georgia.
- MCA-Georgia: Provides project design and legal support as well as seek funding opportunities from the Government of Georgia and private businesses to cover student's (participants') tuition.
- Administrators from SDSU: Design and administer SDSU-Georgia program, engage in recruitment strategies, seek funding and provide support for partner universities to become ready for accreditation of their STEM programs.
- Administrators of partner universities: Facilitate access of SDSU to partner university facilities and faculty, work with SDSU-Georgia on improving the partner universities' STEM programs and achieving accreditation.

- Georgian and American SDSU faculty: SDSU faculty from the campus in San Diego teach in the SDSU-Georgia program either remotely or in person. They also train Georgian faculty, selected competitively, to teach in the same programs.
- Georgian employers: Serve on advisory board to inform labor market needs and provide internship opportunities.

Implementation to date

The SDSU-Georgia program began enrolling students in September 2015. To date, about 530 students are enrolled. The number of students enrolled in the program in the first years was significantly lower than was predicted and desired. This setback energized change and greater intensity in the efforts to recruit students, spread the word about the program, and popularize STEM disciplines. Further, the realization that few Georgian families would be willing to pay higher-than-regular tuition for an unknown program has led to an intensified pursuit of additional private and public funding toward students' scholarships. Fundraising and outreach have become one of the principal engagements and objectives of the Project.

As of December 2018, all programs of study expected to be delivered by SDSU have been instituted across the partner universities, albeit to differing degrees of development. Chemistry and electrical engineering were the first-degree programs offered by SDSU-Georgia, while computer science and construction engineering were the last to join the menu of the SDSU degrees offered in Georgia.

SDSU has taken responsibility for the development and delivery of all its courses, using a mix of SDSU faculty and faculty from the partner universities. The extent of in-person vs. distance instruction has varied across the programs, as have the modes of interaction between the U.S. faculty and Georgian students, with the notion that the in-person presence of the U.S. faculty was particularly important early on in the program and then again for the higher-level courses. The English-speaking Georgian faculty initially played a role of facilitators of the courses within Georgia and lead laboratory sections, conducted Q&A sessions, and helped liaise between the students and the U.S. faculty. Through this process, as well as through additional training opportunities in the United States and Georgia, the Georgian faculty expected to adopt the instruction styles of their American counterparts. It is expected that a Georgian faculty member would facilitate an American faculty member-led course at least twice before teaching it independently. Once the Georgian faculty's expertise was considered sufficient, they began to teach courses independently.

Most partner universities have received significant infrastructure improvements, such as new labs and lab equipment and – in the case of Ilia State University – even a new building. Multiple Georgian students and faculty have visited the SDSU campus in the United States and a number of U.S. faculty and students have now spent time in Georgia as well.

B. Logic Model

The Project is expected to inspire and facilitate improvements to the partner universities' infrastructure and equipment, contribute to local faculty development and implement new curricula, among other activities. These in turn will affect a range of short-, medium-, and long-

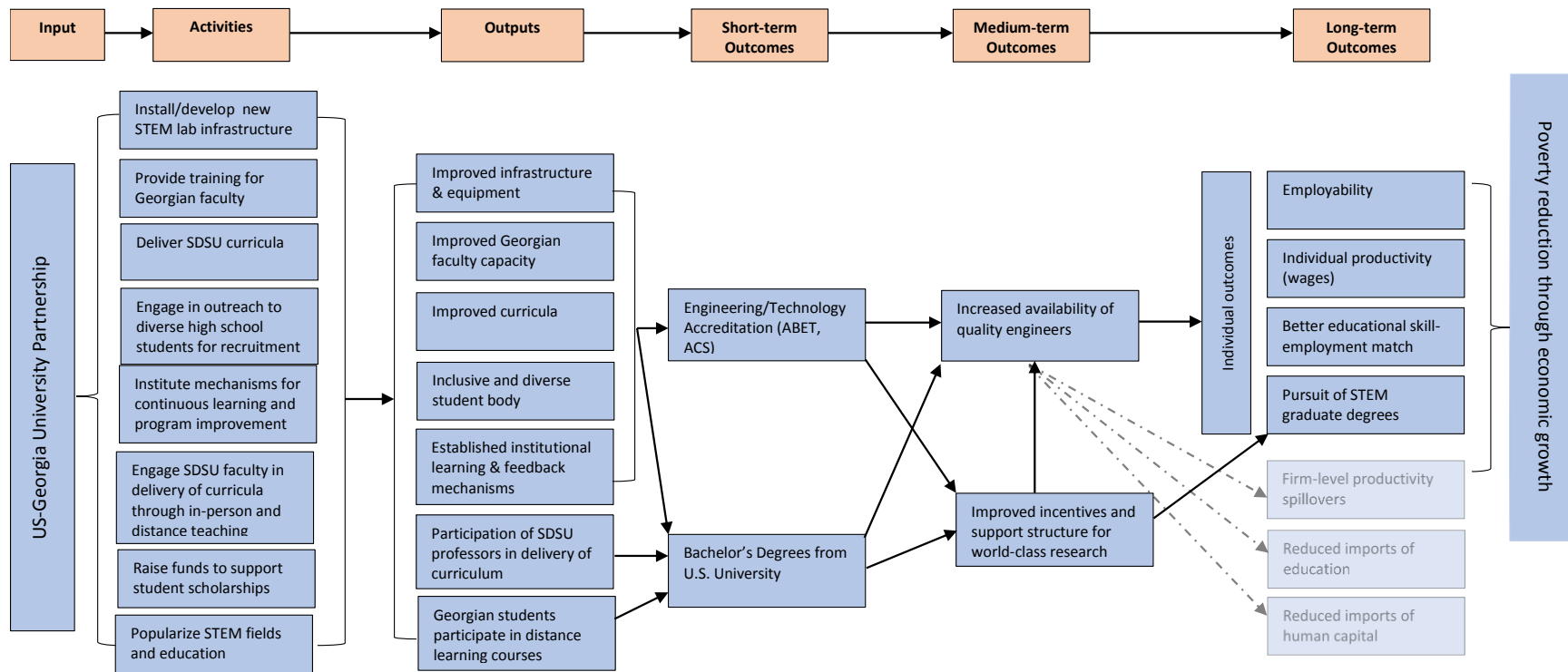
term outcomes. In addition to the activities envisioned in the original logic model,⁶ early enrollment setbacks moved the partners to engage in significant recruitment efforts and popularization of STEM subjects (see Figure 2.1).

According to the logic model, through these activities, in the short term, the Project will facilitate the pursuit and acquisition of engineering and technology accreditations for the participating programs and will grant U.S. bachelor's degrees for the enrolled students. These short-term outcomes would, in turn, lead to an increased availability of quality engineers and improved structures for world-class research in the medium run. In the long run, the program participants are expected to have better employment opportunities and higher incomes than their counterparts who received similar degrees in other programs.

While logic model shows that the Project is expected to ultimately improve household income, provide firm-level productivity spillovers, reduced imports of education (i.e., greater reliance on high-quality educational resources domestically), and reduced imports of human capital (foreign labor), these outcomes will be addressed only in a limited way by the RAND evaluation. The RAND evaluation will emphasize participant outcomes in terms of employability, wages, educational-skill employment match and pursuit of STEM graduate degrees.

⁶ See Monitoring and Evaluation Plan, p. 15.

Figure 2.1. Project Logic Model⁷



Notes:

This logic model was adapted from MCA-Georgia's Monitoring and Evaluation Plan.

The shaded boxes: While Firm-level productivity spillovers and reduced imports of education and human capital are considered desired outcomes of the project, their measurement may be difficult or impossible.

⁷ The Logic Model is adapted from Monitoring and Evaluation Plan (2018), p. 15.

C. Initial Cost-Benefit & Beneficiary Analysis

With a budget of \$30 million for the SDSU-Georgia Partnership, the economic rate of return was estimated to be 10%. This rate of return was estimated using information in the technical proposal and financial proposal received from San Diego State University in February 2014, based on costs and enrollment projections, as well as a variety of documentation and analyses carried out by MCC and the Government of Georgia. As documented in the compact's Monitoring and Evaluation Plan (2018), this return assumed an average operating cost (average annual tuition) of \$7,434 per student in the U.S. degree program and \$1,589 in the ABET-accredited partner programs from Year 7 on. (More recent discussions with the partner universities suggest that the ABET-accredited programs will cost more than this early estimate.) If average annual operating cost/tuition were to rise above \$8,800 per student, the estimated rate of return would decline below 10%.

The initial estimate assumed that the annual student intake for the U.S. degree programs would start at 495 in the first year of the program (compact year 2) and increase to 610 by Year 5. The estimate also assumed that the total annual enrollment in the programs in a given year would reach a steady state of 2,155 starting in Year 7. Over time, the U.S. degree program students would be replaced by the same number of ABET-accredited partner students as the U.S. degree programs are phased out.

However, the actual enrollment for the first year of the program, 2015-16 enrollment (compact year 2), fell short of the estimated 495 students, with only 86 students enrolled. One hundred and twenty-six students enrolled in 2016-17 (Cohort 2), and 201 enrolled in 2017-18 (Cohort 3). The SDSU Strategic Growth Plan (September 2018) estimated the grand total by Year 5 to be 685 students (cumulative) but as of Fall 2018, with Cohorts 1-4, a total of 530 students enrolled. This discrepancy is expected to lower the rate of return on the project considerably.

The economic analysis of the STEM Higher Education Project foresaw four distinct benefit streams supporting the investment:

1. Higher future earnings for graduates of the new programs relative to the amount that these individuals would have earned if they had attended the best Georgian university;
2. Human capital externalities in the form of spillovers from an increase in the supply of well-educated STEM professionals on business productivity and on earnings and wages of other workers;
3. Savings to the Georgian economy from reduced imports of highly educated and more expensive expatriate STEM professionals, and
4. Savings for those students who, in the absence of the new programs, would have pursued more costly undergraduate STEM degrees at U.S. or European universities.

While the Project's initial cost-benefit analysis forecast several benefits to the investment, the scope of the RAND evaluation will focus mainly on examining student outcomes in terms of individual wages, employability, improved matched education-employment skills and pursuit of STEM graduate degrees.

D. Review of Relevant Documentation

This section provides a summary of background documents provided by MCC regarding the Project. They address labor market needs, constraints to economic development and sources of economic challenges, the program design proposed by SDSU as well as evaluation and monitoring plans. Review of these documents provides context for the overall evaluation. The review offers a useful overview of the expectations for the evolution of the project, the predicted challenges, some discussion for how to overcome them, and plans to sustain the gains⁸.

Multi-faceted preparatory work was conducted prior to the onset of the SDSU-Georgia Partnership, to determine the particular needs of the Georgian economy, evaluate constraints that hinder its growth and to identify the gaps and needs in the Georgian labor market.⁹ Two demand studies assessed the extent to which an establishment of a branch of a U.S. university in Georgia would be a useful and sustainable endeavor that would benefit Georgian economy and succeed in the regional market.¹⁰ Another study estimated the income advantage of receiving a dual Georgia-American degree compared to a regular degree from a top Georgia institution.¹¹ The Technical Proposal of SDSU shed light on the thought and planning behind the Project activities, the expectations for the student enrollment, anticipated challenges, and plans for sustainability.¹² The Monitoring and Evaluation Plan also outlined the main assumptions behind the Project and its evaluability.¹³

Using both quantitative and qualitative data, the preparatory studies pointed to low technological innovation¹⁴ and underdeveloped human capital¹⁵ as the primary, binding constraints to Georgia's economic growth, precipitated, in large part, by the low quality of higher education in the areas of science, engineering and technology. According to these analyses, the delivery of high-quality STEM degree programs in Georgia could be a potent way to tackle these challenges.¹⁶

The intervention was designed to address the root causes identified by these preparatory reviews. With the overarching goal of reducing "...poverty through economic growth in Georgia by means of MCC's assistance to strengthen good governance, economic freedom, and

⁸ SDSU Technical Proposal, p. 16-19

⁹ *Constraints Analysis, Georgia, Final Draft*. Government of Georgia, July 2011.

¹⁰ *Demand Study Relating to the Development of Millennium University, Revised Report*. Texas International Education Consortium, May 2011. *Study and Recommendations to the Ministry of Education and Science of Georgia Regarding Establishment of a U.S. University Branch Campus, Final Draft*. Art and Science Group, LLC, June 29, 2010.

¹¹ Hans Gutbrod, CRRC, "Measuring Wealth Differential." Presentation for MCC, 8 November 2011.

¹² *Technical Proposal for Selection of Partner Institution(s) for Capacity Building and Establishing Bachelor Degree Programs in Science, Technology, Engineering and Math Higher Education in Georgia: A Partnership between San Diego State University, Ilia State University, and Tbilisi State University*. San Diego State University, March 2013.

¹³ *Monitoring and Evaluation Plan, Compact II*. Millennium Challenge Account – Georgia, Version 3, January 2018.

¹⁴ see *Constraints Analysis*, pp. 79-87

¹⁵ see *Constraints Analysis*, pp. 32-38

¹⁶ see *Demand Study*, p. 69

investments in Georgia”, the project offers a public good rationale for the necessity of the STEM Higher Education intervention and the accompanying investment.¹⁷ Further, the Monitoring and Evaluation Plan clearly states the Project’s role in strengthening “...the linkage between market-demanded skills and the supply of Georgians with technical skills relevant to the local economy and support delivery of high-quality STEM degree programs in Georgia” (p. 6)¹⁸ through “the long-term delivery of high-quality of STEM bachelor’s degrees in Georgia” (p. 12).

Thus, the logic model guiding the Project has been based on the expectation that the SDSU-Georgia programs, as part of the broader Partnership, would provide high-quality STEM education, which would ultimately lead to the benefits that would spread beyond the degree programs’ immediate graduates. The strength of the evidence for the Project’s theory of change is varied, in part, due to the lack of reliable nationwide data on Georgian STEM graduates and Georgian labor markets (see p. 48). In some cases, when the evidence in Georgia is limited (e.g., on educational spillovers), evidence from other contexts was used (e.g., United States), with appropriate caveats explicitly stated (see p. 25). Still, some assumptions on the benefit streams seem to lack strong empirical basis – e.g., that the intervention would reduce the need to import high-skill STEM workers, as no such data were collected (p. 48).

The SDSU-Georgia Partnership is situated in the broader context of multiple activities working contemporaneously toward the common goal of Georgia’s improved economic performance through improved education. These complementary activities have focused on the improvement of the learning environment infrastructure in schools, professional development for educators, education assessment support, and workforce development through provision of industry-led skills.¹⁹ However, our review did not find a detailed discussion of how the Partnership complements, aligns, or competes with other initiatives in the country – e.g., with the European universities and other foreign-degree programs.

¹⁷ Monitoring and Evaluation Plan, p. 6; Constraints Analysis, p. 32, p. 49; Demand study, p. 3.

¹⁸ In the remainder of this document, references to page numbers are to the Monitoring and Evaluation Plan unless noted otherwise.

¹⁹ See Monitoring and Evaluation Plan, p. 7-12.

3. Evaluation Design

The RAND evaluation addresses five research questions described in more detail below. These questions are primarily derived from MCC's the solicitation for the external evaluation, MCC's logic model presented in Figure 2.1, and discussions with MCC regarding the Partnership, vision and goals. Specifically, the RAND evaluation examines five key areas: (1) implementation of the SDSU-Georgia program and other Project activities; (2) assessment of partnerships between SDSU and public Georgian universities; (3) sustainability; (4) SDSU-Georgia program student outcomes; and (5) economic return on investment.

The success of the Project and, in particular, the success of the SDSU-Georgia program can be understood by examining each of the five areas in depth. For example, before determining whether the SDSU-Georgia program is successful in meeting its targets for student outcomes, it is critical to understand and document how and whether all of its components have been implemented. As previous research on education program implementation has amply documented, the level and quality of implementation is a major determinant of the extent to which the desired outcomes may be realized.²⁰ A related area to whether having program activities were implemented well is the strength of the partnership among SDSU and the public colleges. Strong partnerships will ensure SDSU-Georgia program activities as well as other Project activities (e.g. efforts to support colleges obtain accreditation in STEM programs) are well coordinated among partners and are aligned to achieve the Project's goals.

Sustainability is also critical to assess. Project activities as well as outcomes are likely to suffer if partners do not engage early on in planning for sustaining their Partnership after funding ends. Finally, the SDSU-Georgia program success should not only be measured by its student outcomes, but also by examining its economic return on investment. Programs that are successful but expensive to sustain may not be optimal for improving educational outcomes. Policymakers and education experts might need to reconsider their approach and find more affordable alternatives to improving higher education in Georgia.

Below we present each research question separately and its justification. We will assess each of these areas contemporaneously in the course of five years, using mixed-methods methodologies and drawing data from a wide range of key stakeholders.

A. Evaluation Questions

- Question 1: Were the activities implemented through the Project aligned with the design, as documented in the logic model?

Justification: As indicated in the logic model, there are number of activities in which the partner universities should engage to improve the quality of STEM programs. They include installment and development of new STEM lab infrastructure, providing training for Georgian faculty, delivery of SDSU curriculum, institution of mechanisms for

²⁰ Vernez, G, Karam, R., Marshall, J, School-Based Management in Indonesia, Santa Monica, CA: RAND, 2012.

continuous learning and program improvement, outreach aimed at balancing the gender of program candidates and enrollees, and implementation of distance learning and outreach programs and raising funds for scholarships. Examining whether and how each of these components were implemented is critical to understanding the extent to which the project was implemented as intended. Further, understanding how and why the programs differed in implementation of different components and with different partner universities and student groups (e.g., female and male students, domestic and international students) will help shed light on the potential differences in program outcomes and will inform future efforts.

- Question 2: How was the Partnership established and carried out? How did it change over time?

Justification: At the core of this project is the development of interorganizational partnerships to support educational reform. Specifically, the Partnership between SDSU and Georgian universities is designed to improve STEM education in four-year public colleges. This Partnership is expected to develop a path to accreditation for programs at Georgian partner universities to which the SDSU curriculum is transposed and to support partner universities in obtaining ABET accreditation and chemistry certification (ACS) for a few of their existing Georgian language engineering and computer programs. Examining the nature and strength of this Partnership is critical for informing project implementation (both successes and barriers), outcomes, and future sustainability.

- Question 3: To what extent are Project activities sustainable?

Justification: Central to MCC's investment in developing the Partnership between SDSU and Georgia universities is the sustainability of the program after the Compact closes in 2019. This was highlighted in the solicitation, in relation to finding internal sources of funding for the improvement of STEM programs in three Georgia universities through accreditation and program visibility.

- Question 4: What is the impact of the SDSU-Georgia program on outcomes including income, better skill match to employers, and a greater share of students choosing to pursue graduate education? Does the impact of the program vary by gender, economic backgrounds of students, and the countries that they are from?

Justification: The effectiveness of any project is measured by whether it achieves its goals. The main goal of the SDSU-Georgia program as defined by the logic model is to improve STEM programs for the purpose of developing high quality STEM graduates and improve the employment opportunities and wages of Georgian STEM graduates, and increase the number of Georgian students pursuing graduate education. The SDSU-Georgia program targets females to increase their participation in STEM programs and jobs. The program also includes students from a variety of family economic backgrounds to improve their employability and productivity. Finally, while it is not a primary objective, the SDSU-Georgia is expected to attract international students.

- Question 5: What is the post-compact economic rate of return? How accurate were the original estimates and assumptions?

Justification: MCC has made a large investment for improving the public higher education system in Georgia and the skill of STEM students graduating from these institutions. It becomes critical to compare estimated future benefits from the program to costs incurred in implementing the program to determine whether the partnership approach is economically sound for reaching its outcomes. It is equally important to examine the original estimates and assumptions to learn about the cost-benefit model to inform future projects.

B. Evaluation Design Overview

The evaluation utilizes a mixed-methods approach that involves quantitative analyses of surveys along with document and literature reviews and qualitative analyses of interviews, focus groups and case studies. Such a mixed-methods evaluation will allow us to examine a variety of aspects of the STEM Higher Education Partnership program including partnerships, design and implementation, outcomes, barriers to and facilitators of high-quality implementation, cost, and sustainability.

The study utilizes tracer studies as the primary quantitative approach to examine Project outcomes including student perceptions of the SDSU-Georgia degree programs, employment while enrolled in the university and after graduation, and wages after degree completion. To strengthen our ability to isolate the effects of the Project on student outcomes, we will examine how the outcomes of interest compare between SDSU-Georgia students and other students who attended and completed degrees in similar disciplines at different universities.

We will use a variety of qualitative approaches to gain a detailed understanding of the thinking behind the Project design and the processes that accompanied its implementation. More specifically, we will conduct a detailed Project documentation review and contrast the assumptions that guided the development of project logic with the insights from broader multi-disciplinary literature on similar projects in developing countries. We will then rely on interviews with principal stakeholders and focus groups with faculty and students to gain insight into how the Partnership evolved and the different challenges and developments in the course of implementation. While we will aim to create a wholistic picture of the Project, we will also examine how different processes have evolved in the three public universities. To this end, we will use a range of qualitative methodologies to examine how various stakeholders at different levels of engagement in each of the three partner universities understood, facilitated, implemented, and experienced the Project. Examining each of the three universities separately will enhance the robustness of the overall findings, allow for detection of similar trends across different contexts, and permit identification of barriers to and facilitators of positive outcomes, if the results differ across programs.²¹ Collecting data from multiple sources using a mix of qualitative methodologies will further minimize potential inadequacies in one source of data and will enhance overall validity of the findings. We will also interview MCC, MCA-Georgia and Ministry of Education (current and past staff who were involved and/or have knowledge of The Project) to understand the support provided, policies that hindered or facilitated the Partnership, implementation of the Project activities, and sustainability efforts. Given the high turn in the Ministry staff, we will aim to also interview those that have left their positions. These interviews will also address future policies the Ministry of Education might undertake that affect (positively or negatively) the Project activities and outcomes. From these interviews we will be able to

²¹ Yin, R.K. Case Study Research: Design and Methods. Sage. Thousand Oaks, California, 2003.

identify common themes across stakeholders, partner universities as well as variation in their perception regarding partnership, implementation of activities, sustainability and provided supports.

We summarize the evaluation questions, key outcomes, data sources and types in Table 3.1 and present a timeline for different data collection activities in Table 4.1.

Table 3.1. Evaluation Questions, Key Outcomes, and Data Sources

Evaluation question	Key measures and/or outcomes	Data sources	Data type
Q1. Were the activities implemented through the Project aligned with the program design, as documented in the logic model?	Implementation of activities covering: (1) installing and developing STEM infrastructure and installing equipment; (2) providing training to Georgian faculty; (3) Engaging SDSU U.S. faculty in implementing SDSU curriculum; (4) engaging in continuous improvement efforts; (5) engaging in outreach efforts to attract diverse high school students to enroll in program and popularize STEM; and (6) raising funds for students' scholarships	Current and former MCC staff, MCA Georgia staff,; SDSU leadership and faculty, and TSU, ISU and GTU leadership and faculty	Interviews and focus groups, walk through observation of labs, and document review
Q2. How was the Partnership established and carried out? How did it change over time?	(1) Criteria for identifying a partner organization, (2) partner organization capacity, (3) nature of coordination and collaboration, (4) shared understanding and commitment, (5) shared understanding of what constitutes project success, (6) joint decision making; (7) accountability in pursuit of common outcomes, (8) perceptions of mutual benefit, (9) incentives needed for the success of the program, and (10) presence of communication structures and trust	Current and former MCC staff; MCA Georgia staff; Georgia Ministry of Education; SDSU leadership; and TSU, ISU and GTU leadership	Interviews, Project documentation, and MOU between SDSU and 3 partner universities
Q3. To what extent are the Project activities sustainable?	(1) Popularization of STEM discipline, (2) awareness of, visibility and appreciation, and (3) program accreditation	High school students, Advisory board, Georgia Ministry of Education; and TSU, ISU and GTU leadership and financial information	Interviews, focus groups, and financial and budget reports

Evaluation question	Key measures and/or outcomes	Data sources	Data type
Q4. What is the impact of SDSU-Georgia program on outcomes in income, better skill match to employers and a greater share of students choosing to pursue graduate education? Does the impact of the program differ between males and females, students from different academic background and different countries?	(1) Wages; (2) job characteristics (hours worked, sector worked, occupation, location); (3) employment tenure; (4) skill match with employer; and (5) enrollment in graduate programs.	Students in Partnership programs and students in comparison groups	Tracer surveys
Q5. What is the post-compact economic rate of return? How accurate were the original estimates and assumptions?	(1) Fixed costs of program, (2) on-going costs, and (3) wages	Workplans, budgets, MCC staff; MCA Georgia staff; SDSU leadership; and TSU, ISU and GTU leadership, Georgia government	Review of literature/documents and interviews, tracer surveys, information collected in Qs 1-4.

C. Quantitative Approach

We will utilize quantitative approaches to answer questions 4 and 5 presented in Table 3.1 above. We begin our discussion with Question 4.

Question 4: What is the impact of the SDSU-Georgia program on outcomes in income, better skill match to employers and a greater share of students choosing to pursue graduate education? Does the impact of the program differ between males and females, students from different academic background and different countries?

Design

The quantitative study is intended to survey all students enrolled in SDSU-Georgia as of June 2019 (the first four entering cohorts) and to compare their outcomes against a set of comparable students in other similar programs. Unfortunately, administrative data for other programs is not available for the study to select comparison students. Therefore, we have adopted a two-stage approach to construct the comparison group. We first select programs in the same or similar fields to SDSU-Georgia and those with entrance exam scores as similar as possible to SDSU-Georgia. We then recruit students within these programs to join the comparison group and use propensity score weighting to account for potential differences between SDSU-Georgia and comparison students to the extent possible.

The study will conduct two major waves of surveys of the treatment and comparison groups: a baseline wave in 2019 and an endline wave in 2023. During 2020-2022, the survey company will recontact the sample annually to maintain their current contact information.

Treatment Group

All of the 530 students attending the SDSU-Georgia program as of June 2019 will be included in the evaluation study (“treatment group”). This group includes four cohorts that have entered the program to date. The (first) 2015 cohort will be in their final year at baseline data collection, the (second) 2016 cohort will be in their third year at baseline data collection, and so forth. The average national entrance exam scores for the SDSU-Georgia programs by cohort and major field ranged between 2041 and 2145, which are relatively high compared to many other programs in Georgia.

Comparison Group

To examine the impact of the program, we will construct a comparison group from two different sources: (1) students who are enrolled in the partner universities but pursuing non-SDSU-Georgia programs in the same or similar fields; and (2) students who are enrolled in universities not participating in the SDSU-Georgia STEM Higher Education Partnership but pursuing a degree in the same or similar disciplines as the students in the SDSU-Georgia programs.

The selection of the comparison programs, whether within the partner universities or non-partner universities, will be based on the type of degree programs and admission competitiveness characterized by the average student score on national entrance exams for these programs. All students enrolled in these comparison programs and who are in the same cohort as the SDSU-Georgia students will potentially be included in the comparison group. However, since the SDSU-Georgia program admits on average higher achieving students than other programs, the students participating in the Partnership program are expected to be different from this comparison group on characteristics that are strongly associated with post-graduation outcomes, such as employment and wages. We will account for some of these differences statistically using propensity score weighting (described in more detail below) as we will collect these data on each student in the Partnership program and the comparison group.

The advantage of selecting a comparison group from the universities that are implementing the Partnership program is the reduction of biases resulting from institutional cultural factors. Students attending the same institution are also more likely to be similarly motivated. A disadvantage is the SDSU-Georgia program might spill over to affect the outcomes of the students at the three colleges who are not enrolled in the program, thus reducing our ability to detect the program’s impact. Having the comparison students from other colleges will help address the spillover effect. However, there may be unobservable differences which led these students to choose a different university.

As discussed in the Statistical Power Calculations section below, we will aim to select a comparison group of at least 3600 students. Appendix A presents data on the programs from which we plan to choose the comparison group. If we restrict the comparison group to only programs with average national exam scores equal to or above the minimum of the SDSU-Georgia programs (2041), there are 1,339 students available for the comparison group, which is not sufficient. If we expand the list of the comparison programs to include those with average scores of 1850 or above, there are 9,230 students available for the comparison group. If we lower the score threshold to 1750, the programs in this range could provide up to 13,427 students to the comparison group. The programs with lower scores tend to be larger (some include up to 500 students per year). To avoid skewing the comparison group in favor of a small number of large programs with lower scores, we plan to stop recruiting students from these large programs after a

certain number of students are reached. We will determine the lowest score threshold for the comparison programs and the maximum of number of students to be recruited from each program during the recruitment process based on our progress towards the desired comparison group size of 3,600 students.

Statistical Power Calculations

As mentioned above, the survey data will be collected from a treatment group of current SDSU-Georgia students and a comparison group of students from similar non-SDSU-Georgia programs in partner universities and other universities in Tbilisi with average entrance test scores as similar as possible to the SDSU-Georgia programs.

This study is designed to provide sufficient statistical power to estimate a reasonable effect size on student wages. The treatment group consists of the 530 students who were enrolled in the program through the first four cohorts. Table 3.2 below shows the size of the comparison sample needed to detect a 25 percent increase in wages for the treatment group compared to the comparison group with 80 percent statistical power and Type I error of 5 percent. Below we discuss the different scenarios and the different assumptions that were considered in these calculations.

1. **Average monthly wages:** Gutbrod (2011) used a matched resume study with local employers to estimate that current domestic graduates in Georgia earn a median wage of US\$800 per month and that graduates of an American degree program that have otherwise identical resumes would earn 44 percent more than the median on average. To be conservative, we estimate power based on a 25 percent increase in wages. Gutbrod (2011) estimated that the standard deviation of domestic graduate wages was \$580 if all observations are considered and \$308 if the top and bottom 10 percent of outliers are removed. We used a comparison group wage of \$800 and a monthly standard deviation of \$400 consistent with these figures, which translates the 25 percent increase in wages to an effect size of 0.5 of a standard deviation.
2. **Clustering of individuals within programs:** intraclass correlation coefficients (ICC) ranging between 0 and 0.1. Larger ICC values indicate more correlation across individuals within programs and universities, making it harder to detect differences across programs and therefore lowering power. We assumed an average of 50 students per cluster. We do not know the extent to which covariates that will be included in the models will be able to explain the between cluster variance. An ICC of 0 corresponds to the optimistic scenario where all between cluster variance can be explained with the covariates.
3. **Response rate at baseline in Year 1:** For comparison students, the response rate is 100 percent by design. We are aiming to survey at least 90 percent of treatment students, which may be ambitious, but SDSU has agreed to cooperate closely to maximize the response rate among treatment students. Overall, we assume that 90 percent of the both the treatment and comparison group provide usable responses, to account for some non-response among treatment students and key item non-response among both treatment and comparison group.
4. **Response rate at the final follow-up in Year 5:** Other tracer studies have achieved response rates of about 30% when following up multiple years after graduation with no efforts to prepare or maintain contact with respondents for the survey. Since we expect to

maintain contact with respondents over time, we hope to achieve a better response rate, especially in the treatment group which is likely to remain engaged with SDSU and the project. We therefore simulate several assumptions of final follow-up rates in Year 5 of 30%, 45%, or 60% for the comparison group. The response rate for the treatment group is assumed to always be 15% higher than the comparison group (45%, 60%, and 75%).

5. **Differences in characteristics between the treatment and comparison group:** As described in detail below, we will use a propensity score weighting approach to account for the differences in the characteristics of the treatment and comparison students. We assume that this adjustment will lead to a design effect of 3, i.e., the effective size of the analytic sample size is assumed to be one-third of the total sample. We do not know the extent to which characteristics of the students will be able to explain student-level variances; therefore, we did not account for this in the power calculations, and the calculation might be conservative.
6. **Comparison group size:** roughly three to eight times the size of the treatment group (1800, 2400, 3000, 3600, 4200, and 4800).

For the most conservative scenario a comparison group of 3600 students is required to meet the 80% power threshold. For optimistic scenarios, a comparison group of 2400 appears sufficient. If the actual wage differential is below 25 percent, we may have difficulty detecting it, although we are making fairly conservative assumptions and may have some additional power beyond these estimates.

Table 3.2. Power to detect a difference in monthly wages between 800 (comparison) and 1000 (treatment)

Scenario				Power by Comparison Group Size					
Type	ICC	Treatment follow-up	Comparison follow-up	1800	2400	3000	3600	4200	4800
Conservative	0.10	45%	30%	48%	65%	77%	84%	88%	91%
Moderate	0.10	60%	45%	62%	78%	87%	92%	94%	96%
Optimistic	0.10	75%	60%	72%	85%	92%	95%	96%	98%
Optimistic	0.05	45%	30%	71%	86%	93%	96%	98%	99%

Primary Data Collection

We will utilize a combination of face-to-face, electronic, and phone surveys to collect information from students during the study. Both the treatment and the comparison students will be surveyed twice, in year 1 (the baseline) and at the end of the fifth year of the study (the endline). In years 2, 3 and 4 of the study, we will implement short follow up phone surveys to sustain interaction with students and ensure that their contact information is up to date.

The baseline survey will be administered using face-to-face, electronic, and phone modalities in the first year of the study. All currently enrolled students from each cohort (and comparison groups) will be surveyed to gather a range of detailed information, including demographics and

socio-economic background, national test scores, information about their overall course of study, perceptions, opinions of and experiences in their degree program, participation and internships, their plans and aspirations after the graduation, including whether they plan to remain in Georgia or pursue jobs and education abroad.

The endline survey will be administered electronically or via phone in the fifth year of the study and will aim to learn about SDSU-Georgia graduates' experiences after graduation. The endline survey will inquire about graduates' perceptions of and experiences in the SDSU-Georgia program, their labor market and employment outcomes. Specifically, the survey will gather information on employment status, hours worked, earnings, job location and characteristics, wages and benefits, and employment tenure and likelihood of retention. The survey will also capture the circumstances leading up to employment such as time to find a job after graduation and approaches the students took to find their current job. In addition, the survey will assess the graduates' perceptions of their employment, such as the match between the job and their interests, skills, and educational background, overall job satisfaction, and the extent to which graduates deemed their education and training had prepared them for the job. We will also collect information on whether students pursued graduate education, in which university, and why.

For those not working, we will gather information on how actively they have been seeking work, by assessing such factors as the specific job search activities, including the number and types of jobs to which they applied, number of interviews, and job offers. The survey will also inquire about all graduates' perceived barriers to finding work.

Conducting only two principal waves of data collection imposes limitations for the range of the experiences the evaluation will be able to capture, as well as limits the timeline for the program impact to fully flourish for the SDSU-Georgia students in later cohorts. Funding restrictions prevent us from conducting a greater number of survey waves between Years 1 and 5 or from following up at a later point. However, while not without caveats, surveying SDSU-Georgia students from different cohorts will allow us to track (and, if we have sufficient power, compare) their outcomes at different points after their graduation (immediately after graduation for the cohort that started in 2018 and four years after the graduation for the cohort that started in 2015) based on students graduating in four years, which appears typical in Georgia. In addition to the survey, in year 5 we will interview 15 employers who have hired graduates from the SDSU-Georgia program to understand about their performance in their workplace and whether their skills and preparation match employer needs.

Piloting the Survey and Entering Data

MCA-Georgia has identified Analysis and Consulting Team (ACT) – a global research-based consulting team headquartered in Tbilisi, Georgia – as the contractor for the data collection. Since the selection, RAND has worked closely with ACT representatives to identify the most appropriate methods of data collection. RAND has also worked with ACT to pilot the questionnaire with Georgian students and ensured that necessary adjustments are made and reflected in the Georgian-language versions of the survey.

Further, RAND has worked together with ACT to define procedures for data quality control during the main administration such as : (1) monitoring the work of the survey interviewers through back-checks including real time observations during the first week and subsequently as needed; (2) reviewing the questionnaire responses to identify if questions are skipped, answers are ambiguous, or identify other problems with questionnaires; and (3) provision of re-training of

enumerators or mitigating difficulties if systematic problems are found. RAND and ACT collaborated in training the key staff, including enumerators and supervisors, who will be involved in the data collection. ACT has been in touch with the participating universities to establish best suited approaches to data collection.

RAND has also worked closely with ACT to ensure data safety and minimal data entry errors. ACT will be responsible for utilizing a data entry software that allows for data checks to be programmed into the software to maximize data quality. For surveys administered on paper, ACT will double-enter the data, compare them, and reconcile any differences. During and after data entry, ACT will check the entered data for logical inconsistencies and return to the original questionnaires to resolve them. If any inconsistencies cannot be resolved by reviewing the original completed questionnaires, ACT will consider contacting the respondents to resolve the inconsistencies. If it is not possible to recontact the respondents, missing values will be used consistently. RAND will also be requesting periodic “data dumps” for additional quality checks on the data.

All changes made to the originally recorded questionnaire data will be fully documented, and a file preserving the original uncorrected data. ACT will also ensure that all variables are named and labeled according to specifications provided by RAND. Once the survey data are cleaned and labeled, ACT will provide a codebook that will include a description of all variables collected.

As the surveys will only be implemented in the first and fifth year of the evaluation, respondent attrition might be high. RAND will work closely with ACT to ensure adequate response rate. As mentioned, ACT will sustain continuous contact with participants by contacting students in-between survey years to ensure continued interaction and updated contact information. ACT will also send out multiple reminders as a standard operations procedure to increase the response rate. Finally, to increase response rate all surveys are (or will be) designed to be completed in less than 30 minutes.

Analysis Plan

We will develop empirical models that incorporate the outcome variable as the dependent variable and the factors that are known to affect that outcome as the covariates or independent variables. Participation in SDSU-Georgia program will be included as the principal predictor of the outcomes of interests (whether students participated in the program or are in the comparison group).

Despite our attempts to select a comparison group that is similar to the treatment group, the observable characteristics of SDSU-Georgia students are likely to differ from the characteristics of the comparison students. We will implement a propensity score weighting approach to reduce bias due to these differences. This weighting is intended to make the weighted comparison group appear as similar to the treatment group as possible. It is important to note that biases due to unobservable factors that may affect student outcomes may still be present.

Specifically, let A_{ij} be an indicator of being in the treatment group for student i in program j , and let X_{ij} be a set of baseline characteristics for the same student. We will use the following logistic regression specification to estimate the propensity score:

$$\text{logit}(\Pr(A_{ij} = 1)) = \alpha_0 + \alpha_1 X_{ij}$$

The set of baseline characteristics X_{ij} will include indicators of cohorts, student demographics, and national test scores. A propensity score weight for each student will be constructed as 1 for the treatment students and $\frac{p}{1-p}$ for the comparison students, where p is the predicted probability from the propensity score model. The quality of the propensity score weights will be assessed by comparing the characteristics of the treatment students to those of the weighted comparison students. If balance is not achieved between the groups a more flexible propensity score model will be used, such as one that optimizes balance using a generalized boosted model.²²

After the estimation of propensity scores, the treatment effect on the outcomes will be assessed while taking into account that (1) the cohorts have different amount of time since graduation and follow-up and (2) the nesting of students within programs. Using wage as an example: let Y_{ij} denote the wage of student i in program j , and let T_{ij} be the time since graduation for the same student. A propensity score weighted version of the following model will be estimated:

$$Y_{ij} = \beta_0 + \beta_1 X_{ij} + \beta_2 A_{ij} + g(T_{ij}|\beta_3) + A_{ij}g(T_{ij}|\beta_4) + \gamma_j + \epsilon_{ij}$$

where $g()$ is a function that captures the relationship between time since graduation and wages, γ_j is a program-level random effect accounting for nesting of students within programs and ϵ_{ij} is a mean zero error term. The coefficient β_2 is of primary interest, which measures the immediate effect of the treatment on wages after graduation. Also of interest is the function $g(t|\beta_4)$, which estimates how the effect of the treatment changes for each year after graduation. Possible choices of $g()$ include indicators for the number of years since graduation, a linear effect of time since graduation ($g(T_{ij}|\beta) = \beta T_{ij}$), or no time effect ($g(T_{ij}|\beta) = 0$).

Since not all students may graduate, graduate on time, or graduate from the same program they are measured in at baseline, we will also explore such potential effects using the data, such as stratifying by cohort, although power to detect differences is likely to be limited. If appropriate we will consider alternative specifications that incorporate such effects.

Other analyses will follow the same general strategy, but with minor modifications to make the models appropriate for the context. For example, when estimating the treatment effect on employment or pursuing graduate degree, the outcome model for employment or graduate degree will be a logistic regression, but all other features will remain the same.

To understand if the treatment effect varies across subgroups (e.g. is the treatment more effective for males than females or for domestic students than international students?), a new set of propensity score weights must be derived that ensures the characteristics of the treatment and comparison students are similar within the subgroups. These weights will be derived by fitting a similar propensity score model as previously described, but it will incorporate interactions between the subgroups and the baseline characteristics X_{ij} . For the outcome model, interactions between the subgroups and treatment will be included. These interactions will estimate how much the treatment effect varies across subgroups.

²² McCaffrey, Daniel F., Greg Ridgeway, and Andrew R. Morral. "Propensity score estimation with boosted regression for evaluating causal effects in observational studies." *Psychological methods* 9.4 (2004): 403.

Challenges and Limitations

A principal challenge will be meeting the desired comparison group size and recruiting to the comparison group students who are as comparable as possible to the treatment students. If we cannot recruit the targeted number of comparison students, the estimation approach that uses propensity score weighting may not have sufficient statistical power to detect effects (even though the power analysis used conservative assumptions). In this case, we plan to estimate impacts using a simple unweighted multiple regression approach, which will still control for student characteristics that may affect student outcomes in both the treatment and control groups. This approach is less preferred than the propensity score weighing approach if sufficient power is available for the latter, but the former could provide valuable information if statistical power is limited for the latter.

Estimating Rate of Return

The treatment effect estimates on wages represent a key input in the rate of return calculations for Question 5.

Question 5. What is the post-compact economic rate of return? How accurate were the original estimates and assumptions?

In order to estimate the post-compact economic rate of return, the team will first identify key input, output, and outcome measures (e.g., program costs and wages) through document review and stakeholder interviews. For each measure, the team will identify the source of data and method of collection, if required. The approach will be to compare estimated future benefits from the program to costs incurred in implementing the program.

In terms of program costs, it will be important to delineate both the fixed costs of designing and implementing the program and the ongoing costs associated with implementation. These will include costs associated with new infrastructure upgrades and investments, increasing capacity of existing faculty and staff through training as well as the need for newly hired staff, curriculum development, student recruitment, and any other costs specifically associated with the Partnership program. Official historical budget documents from SDSU and MCA-G are expected to provide most of the information on costs, with additional information from future budget projections and implementation planning documents that extend beyond the evaluation period. Any additional costs borne through implementation including the value of any in-kind contributions, volunteer time, and other new costs borne as a result of the program and that is not accounted for in the Partnership documentation will need to be detailed through additional document review and discussions with stakeholders. When reviewing the data sources, it will be important to ensure there is no double counting of costs.

Finally, all costs will need to be normalized to the same currency using the most current Georgian Lari-USD exchange rate. This data may be available via official government resources or other macroeconomic resources. Ideally, existing projections will be used. Extrapolations based on recent historical data may also be used. It may be necessary to assume a steady exchange rate over the entire period of analysis.

In terms of estimating benefits, the expectation is that the key benefit to participating in the program is the increase in educational attainment, employment, and total earnings (wages and benefits). Beyond individual-level benefits, there is also a potential benefit to reducing skills gaps in STEM occupations that accrue to both employers in Georgia and society as a whole to

the extent that better matches between supply of skills and demand for skills lead to growth in business and ultimately the economy as a whole. The feasibility of incorporating these types of benefits is unknown without further discussions with MCC and local stakeholders. As the research team sees it, the main challenge with estimating benefits is that these tend to accrue to individuals and society over a long period of time. Thus, the full benefits of participating the program, particularly relative to the initial costs of designing and implementing the program, will not be fully realized during the time period allotted to the study. It may be possible to estimate long-term benefits to graduates of comparable U.S. and other international programs if those subgroups can be identified in extant labor force or other data. A key assumption of this approach is that the SDSU-Georgian University STEM Higher Education Partnership program is comparable to the education and training that Georgian students who study in the same fields at comparable universities abroad receive.

To estimate wage differential benefits over the long run, the following key assumptions will need to be made:

1. Number of years to extrapolate to adequately estimate economic rate of return (ERR): In the initial MCC ERR analysis, a time period of 30 years post-graduation was used. This assumption seems adequate to calculate an ERR.
2. Current annual total compensation adjusted for probability of employment for graduates of U.S. based STEM program: There are multiple publicly available sources for wage data, however, most of these sources don't have detailed data for specific types of jobs, particularly STEM based careers. The best source of data will come from surveys of individuals or hiring firms for wages in STEM fields. Additionally, we ideally would like to capture total compensation (i.e. wages plus other benefits), not just wages. Accurately including such detail will likely require data gathered from surveys.
3. Current annual total compensation adjusted for probability of employment for graduates of local based STEM program: As with annual compensation for graduates of the U.S. based program, total compensation data for those graduating from local STEM programs should be gathered via surveys of recent graduates and hiring firms in STEM fields.
4. Projected real income growth in Georgia: Projections of real per capita GDP can be used to project real income growth. There are likely publicly available resources that report these data, or it may be necessary to extrapolate based on recent historical per capita GDP statistics either reported via official government resources or other macroeconomic analysis resources. It may be necessary to assume a steady real income growth rate over the entire period of analysis.
5. Projected employment rate: Similar to projected real income growth statistics, this data may be available via official government resources or other macroeconomic resources. Ideally, existing projections will be used, or extrapolations based on recent historical data may be used. It may be necessary to assume a steady employment rate over the entire period of analysis.
6. Projected Georgian Lari-USD exchange rate: Similar to projected real income growth statistics, this data may be available via official government resources or other macroeconomic resources. Ideally, existing projections will be used, or extrapolations based on recent historical data may be used. It may be necessary to assume a steady exchange rate over the entire period of analysis.

One likely complication in estimating wage differentials is that SDSU graduates may be more likely to pursue graduate education than comparison students and hence delay the start of their career earnings. While our survey will capture indications of graduate degree enrollment during the evaluation period, it may miss the start of the career earnings for these students.

Our survey will examine the wage differentials between SDSU graduates and comparison students. Following the anticipated end of the SDSU degree programs in Georgia, partner universities may be offering accredited programs which could prepare students for higher wages. We will use available information at the time of the evaluation to estimate the number of students that are expected to complete such programs over a period of years following the SDSU program. Through discussions with employers, we will attempt to estimate how wages for such students may compare to SDSU graduates as a guide to the wage differentials that may be experienced in future cohorts within the ERR analysis. We will also include the relevant income tax revenues associated with wages.

While wage differences will be the largest potential benefit stream, there are other potential benefits to consider. For instance, the MCC ERR included benefit streams for the following:

- human capital externalities in the form of spillovers from an increase in the supply of well-educated STEM professionals on business productivity and on earnings and wages of other workers;
- savings to the Georgian economy from reduced imports of highly educated and more expensive expatriate STEM professionals, and
- savings for those students who, in the absence of the new programs, would have pursued more costly undergraduate STEM degrees at U.S. or European universities.

Estimating human capital externalities is extremely challenging and requires strong assumptions which may be difficult to support with evidence. The underlying rationale behind the benefit stream is logical, however, data existing to estimate the impact of these externalities is scant. The MCC ERR analysis used several U.S. studies to estimate the growth in wages for individuals based on a one percent increase in the share of college graduates. The MCC analysis does admit that studies based on the U.S. economy may not be applicable to the Georgian economy but a lack of studies for the Georgian economy or other similar economies was unavailable at the time of the analysis and may not become available during the timeframe of this evaluation. Further analysis is required to determine whether this impact should be captured and what the estimated impact should be.

The other benefit streams may only be addressed through qualitative discussion. The Georgian economy could benefit more generally from reduced imports of foreign STEM professionals, but we do not anticipate finding quantitative sources to estimate these benefits. We will discuss this potential labor substitution with employers through interviews.

D. Qualitative approach

We chose a qualitative case-study analysis approach to answer questions 1 (implementation), 2 (partnership), and 3 (sustainability) for two reasons. First, the implementation of activities and evolution and sustainability of the Partnership are complex and unfolding, and we had no ability to control the fluid situation. These questions are primarily concerned with description of progress and how progress is made, lending themselves to qualitative description. Second, the complexity of the implementation of the Project created more variables of interest than could be supported by a quantitative analysis.

Below, we present each question and discuss the approach we will take to answer the question. Table 4.1 presented in the next chapter summarizes each major research question, several subtopics for each, and the sources of information and years from which the data will be collected and number of interviews/focus groups. The type of data collection (interviews or focus groups) are primarily determined by the type of stakeholders and information obtained. For example, focus groups are best conducted with faculty groups and student groups as the questions mostly address their experiences with the program. One-on-one interviews are best conducted with leadership as some of the interviews might address sensitive topics regarding barriers and facilitator of implementation, nature of partnership and sustainability of the activities.

Between October 2018 and March 2019, we collected our initial round of qualitative data from current and former MCC staff and consultants, SDSU leadership and faculty, MCA-G staff, Georgian partner universities, employers, and Georgian government officials.

Question 1. Were the activities implemented through the Project aligned with the program design, as documented in the logic model?

We used the logic model, SDSU proposal, conversations with MCC and MCA-Georgia stakeholders, as well as interviews already conducted with SDSU leaders to identify the primary components, processes, and outputs considered essential for the attainment of the program outcomes how they were expected to be implemented. Project activities include: (1) developing and installing infrastructure and equipment (i.e. development of facilities considered essential for the implementation of the program, purchasing of equipment); (2) providing capacity building (e.g., Georgian faculty observation of how SDSU teach the course); (3) engage SDSU-U.S. faculty in delivering SDSU curriculum; (4) institute learning and feedback mechanisms for continuous improvement (i.e., established framework for the improvement of program practices); (5) engage in outreach to diverse high school students for recruitment (e.g., activities designed to promote awareness and accessibility of the program) and popularize STEM; and (6) raise funds to support students' scholarships.

While all SDSU-Georgia degree programs followed a similar general approach to program implementation (e.g., all courses are delivered by SDSU-Georgia faculty, labs updated), no specific guidance was agreed upon on how activities had to be implemented. For example, no uniform guidance across programs was issued on how much of each course was to be taught by the U.S. faculty in distance or in person. Since most of the activities did not have to follow specific prescriptive practices, RAND created a checklist for the types of activities that were expected to be implemented by the Project, and, in Year 1 of data collection, has relied on walk throughs, interviews, focus groups, and document reviews to understand how the activities were implemented within SDSU-Georgia program and identify what worked, what adjustment needed to be made and what the barriers and facilitators to implementation were.²³ We also looked at whether the services provided or implemented such as recruitment strategies varied by student gender. We examined barriers and facilitators to implementation and roles MCC, MCA-Georgia and the Ministry of Education had to support implementation (e.g., obtain regulatory approvals and licenses to enroll international students). We collected this data in year 1 and will continue collecting similar data in years 2,3,4, and 5 to examine changes in implementation.

²³ Tracer surveys – a quantitative approach – also incorporated questions relevant to implementation.

Question 2. How was the Partnership established and carried out? How did it change over time?

One of the evaluation objectives is to examine the nature of the Partnership between SDSU and the three Georgian universities, all together and with each one individually. Partnerships are challenging and often develop slowly and unevenly. Partnerships also demand new behaviors among all the partners, often requiring participating actors to make significant changes and resource investments. Understanding the capacities of participating organizations to make change and what may be required to bridge gaps is critical to uncovering the barriers to and facilitators of the implementation of the Project. Further, the nature of the collaboration among partners is likely to be affected by such factors as their perceptions of the need for the program, understanding of what constitutes program success and how to attain it, and inter-personal and inter-organizational relationships.

In year 1, we reviewed the literature on effective partnerships in international settings and on STEM initiatives in higher education. Based on this review, we identified best practices guided the development of the data collection instruments in general and the Partnership evaluation specifically. In year 1, we also assessed how the Partnership developed and its current status. Using the partnership best practices, we examined such elements as: (1) MCC, MCA-G and SDSU's approaches to the identification of partners and criteria for partner selection; (2) whether the partners were the proper partners and whether key ones were missing; (3) nature of coordination and collaboration among the partners; (4) shared understanding of and commitment²⁴ to the program vision and partnership goals; (5) shared understanding of what constitutes project success; (6) joint decision making; (7) accountability in pursuit of common outcomes and development of joint products; (8) perceptions of mutual benefit by the partners ("win-win model"); (9) incentives needed for the success of the program; (10) presence of clear relationship management guidelines and communication structures and frameworks and; (11) mutual trust (contractual, competence, and goodwill).²⁵ In years 2 through 5, we will also assess how the partnerships changed over time in terms of membership and the mentioned elements above.

To conduct the assessment of the Partnership, we have sought input from a number of key stakeholders, including (but not limited to) the ones on the list provided by MCC.²⁶ We further expanded this list to ensure that all relevant actors are included: program funders (MCC and MCA-G), relevant administration officials of the partner universities both in Georgia and the United States, U.S. and Georgian actors in charge of moving the programs toward international accreditation across all participating universities, and employers (all summarized in Table 4.1. below). We assessed all areas of interest to the Partnership in year 1 of the evaluation.

²⁴ *Commitment* may be defined as "a desire to develop a stable relationship, a willingness to make short-term sacrifices to maintain the relationship, and a confidence in the stability of the relationship" (Anderson, E. and Weitz, B. "The use of pledges to build and sustain commitment in distribution channels." *Journal of Marketing Research*, 1992, vol. 29: 18-34). Morgan and Hunt identified the need for commitment between partners as "key to achieving valuable outcomes for themselves" (Morgan, R. and Hunt, S. "The commitment-trust theory of relationship marketing." *Journal of Marketing*. 1994, vol. 58, no.3: .20-38).

²⁵ Sako notes that trust may be manifested in three ways. Contractual trust refers to the parties adhering to written or verbal contracts. Competence trust refers to the expectation that a partner can perform at a set level. Goodwill trust is the demonstration and mutual expectation of honesty and benevolence between the parties (Sako, M. Price, Quality and Trust: Inter-Firm Relationships in Britain and Japan. Cambridge University Press Cambridge, 1992).

²⁶ See p. 16 of Solicitation

We will continue to assess the Partnership annually in year 2 through 5 (specifically items 3 to 11 above), to identify any changes within the Partnership that may have an impact on the Project outcomes (especially after the compact closes in 2019).

Question 3. To what extent are the Project activities sustainable?

To answer this question in year 2 through 5 we will look at: (1) popularization of STEM disciplines; (2) improved STEM education offered by the partner universities through obtaining accreditation for their STEM programs; and (3) indicators of financial sustainability. For a program to be sustainable, it should be visible and viewed by stakeholders as legitimate and reputable, benefiting students and the economy. The program should also be affordable by families while at the same generating revenues to the colleges.

We will assess STEM awareness in two ways. First, we selected 10 high schools within and beyond Tbilisi to examine the extent to which high school students from schools with STEM emphasis are aware of the STEM programs in partners universities and future accreditation tracks in the partner universities), students' perceptions of the quality of such programs, whether they are considering careers in STEM fields, and whether they would consider accredited partner programs for their college education. The high schools were selected purposively to reflect the characteristics of students who might consider the SDSU-Georgia program or accredited STEM from the three public universities. While purposeful, non-representative selection of high schools may not be ideal for drawing generalizable conclusions about awareness and perceptions of the programs, under resource constraints this approach will be able to shed light on whether popularization and outreach efforts had reached the most obvious target audiences. To ensure that we capture students from different demographic backgrounds, we will incorporate both public and private schools into the sample. In each high school, we will conduct one student focus group consisting of 8-10 students (of which about half are females) enrolled in the final grade of high school.

Second, we also are selecting a sample of approximately 15 relevant employers from a larger list identified by MCC and MCA-Georgia and assessing their awareness of the Project in a series of structured interviews. The interviews incorporate questions about their awareness upcoming partner university accreditations and how important they consider these educational endeavors for their industries and businesses.

Finally, we will examine indicators of financial sustainability. Specifically, we will interview the partner universities regarding their tuition setting for their accredited programs, whether they are generating sufficient revenue (or are likely to in the future) and whether they are affordable to families. We will also interview Ministry officials to understand whether they have implemented policies that allow universities to set different levels of tuition, especially for internationally-accredited programs, and what level of public tuition scholarships they are making available to families.

Data on sustainability will be collected in years 1 through 5. By collecting this information longitudinally, we will be able to capture changes in program awareness and legitimacy as well as assess partner sustainability efforts through obtaining accreditation to their STEM programs.

Primary Data Collection

As indicated under each research question above, data will be collected primarily through interviews, focus groups, or review of documents. We also indicate above the timing of data

collection for each research question. RAND will primarily be responsible for conducting the interview and focus groups with the Georgian government and MCA-Georgia. RAND sub-contracted with Education Strategy Center to assist in collecting data from employers and Georgian universities. RAND will be responsible for overseeing the quality of Education Strategy Center's data collection. RAND will train the Education Strategy Center staff, who are listed in Chapter 5. All interview data collected will be translated into English. The RAND team will read the transcripts to ensure that the Education Strategy Center did not deviate from implementing the interview protocol and have not asked questions in a way that are leading. If deviation is found, additional training will be provided.

Analysis Plan

All Georgian interview notes will be translated into English. A sample of the translated interviews will be reviewed by research team members against Georgian to ensure accuracy. Content analysis will be conducted on the English version of the notes by two team research members separately to validate results.

For Q1 (implementation), since the Project activities do not define prescribed practices, the interview responses will be organized under each of the six activities. Two team members will review the text under each activity and develop additional sub-codes to identify practices for each activity. The interview data will then be analyzed for patterns across different stakeholder groups to identify common themes and lessons learned and identify areas of divergence. In the analysis, we will also look for themes addressing gender issues. For example, we will look specifically at whether some of the outreach practices targeted women or students from lower economic status or international students for the purpose of increasing their representation in the STEM programs. We will also examine factors hindering or facilitating implementation and support provided by MCC, MCA-Georgia and Ministry of Education/Government of Georgia.

For Q2 (partnership), we will create specific indicators from the literature on the eleven key areas of best practice in developing strong partnerships. We will then identify practices implemented by the Partnership from the interviews and organize the identified practices within each of the eleven key areas identified in the literature. We will compare the practices to understand extent to which the Partnership established between SDSU and the three public universities incorporated best practices.

Finally for Q3 (sustainability), we will organize interview responses across three dimensions of sustainability mentioned earlier and identify the practices implemented by SDSU-Georgia and the three public universities. The research team will review the text and identify the specific practices that were adopted to increase visibility, popularity and improve current STEM programs and identify whether any practices targeted women or other student groups. We will examine the practices within each of the three Georgian partner universities as well as across these universities.

The research team will use commercial products such as Atlas, NVivo or Dedoose to code the data.

4. Data Collection Summary and Timeline

Table 4.1. Topics, Sample, and Timeline for Stakeholder Engagement

	Interviews with SDSU-G leadership	Interviews with SDSU-U.S. leadership	Interviews with SDSU-U.S. faculty	Interviews with partner univ. leadership	Focus Groups with Georgian faculty	Focus Groups with SDSU students	Focus Group with HS students	Interviews with Ministry of Education	Interviews with MCC	Interviews with MCA-G
Sample	2 indiv.	1-3 indiv.	3-4 indiv.	3 groups	3 groups	2 groups	10 groups	1 indiv.	1-5 indiv., as available	1-5 indiv., as available
Years	1-5	1-5	1-5	1-5	1-5	1-5	1 & 5	1-5	1	1-5
Q1: Activities										
Infrastructure	X	X	X	X	X	X		X	X	X
Capacity building	X	X	X	X	X	X		X	X	X
Outreach to students	X	X		X	X	X	X	X	X	X
Q2: Partnership										
Selection of partners	X	X	X					X	X	X
Collaboration	X	X	X	X	X				X	X
Shared understanding	X	X	X	X	X				X	X
Q3: Sustainability										
Popularization of STEM	X	X	X	X	X	X	X	X	X	X
Awareness of SDSU	X	X	X	X	X	X	X	X	X	X
Accredited programs	X	X	X	X	X	X	X	X	X	X
Q4: Impact										
Earnings										
Skill match to employers	X	X	X	X	X	X		X	X	X
Q5: Economic return										
Rate of return										
Validity of assumptions	X	X		X				X	X	X

Table 4.1 Continued

	Interview with Others	Interviews with employers	Tracer Surveys	Contact surveys	Documen- tation & literature review	Walk Through	Budget documentat ion review
Sample	1 Center for International Education	15 indiv. or groups	2800- 5400	2800- 5400	Documents	3 partner univs.	Documents
Years	1	1-5	1 & 5	2, 3, 4	1-5	1-5	1-5
Q1: Activities							
Infrastructure			X		X	X	
Capacity building			X		X		
Outreach to students	X		X		X		
Q2: Partnership							
Selection of partners		X			X		
Collaboration					X		
Shared understanding					X		
Q3: Sustainability							
Popularization of STEM		X					
Awareness of SDSU		X					
Accredited programs		X			X		X
Q4: Impact							
Earnings		X	X				
Skill match to employers		X	X				
Q5: Economic return							
Rate of return		X	X				X
Validity of assumptions		X					X

5. Administrative Information

Human Subjects Protection

RAND's Institutional Review Board (IRB), the Human Subjects Protection Committee (HSPC), has oversight of the entire evaluation study. RAND HSPC has approved the procedures to collect and store data from all procedures in the project.

All project activities will also be conducted under the applicable Georgian law protecting human subjects.

Data will generally be stored in a de-identified form with identifiers restricted to crosswalk files that will be carefully controlled. In the case of crosswalk files stored by ACT, these will be stored on a server isolated from the internet and backed up and shared with RAND only through physical media. Other files not containing identifiers will be stored on network-accessible computers and shared with RAND through a secure file transfer system.

Preparing Data Files for Access, Privacy and Documentation

A future version of this evaluation plan will detail plans for preparing data files and documentation that can be used by other researchers once the project concludes. Procedures will respect the privacy of individual respondents.

Dissemination

Both baseline and final evaluation reports are planned with public dissemination for each.

Evaluation Team Roles and Responsibilities

RAND will collaborate with the Education Strategy Center in Georgia to conduct this evaluation. ACT Global, a Georgian firm, has been appointed as the local data collection contractor by MCA-G.

Dr. Charles A. Goldman will lead the team as Program Manager. Dr. Goldman, an economist, has 25 years of experience conducting mixed methods evaluations of tertiary education programs around the world. Dr. Goldman will also oversee all quantitative aspects of the evaluation.

Dr. Goldman will be assisted throughout the project by **Dr. Rita Karam** as Deputy Program Manager. Dr. Karam, an education policy researcher, has 16 years of experience conducting mixed methods evaluations of primary, secondary, and tertiary education programs around the world. Dr. Karam will also oversee all methods for interviews, focus groups, site visits, document review, and case studies.

Drs. Goldman and Karam will manage a team of US- and Georgia-based experts who will collaborate on all facets of the project.

Dr. Katya Migacheva, a social psychologist, will support the entire evaluation, focusing on interviewing, instrument development, and document review.

Dr. Matthew Cefalu, a statistician, will take lead responsibility for sampling design and analysis for the tracer study.

Mr. Thomas Goughnour, a cost analyst, will participate in the analysis of economic costs of benefits.

Dr. Troy Smith, an economist, and **Dr. Louay Constant**, an education policy researcher, will support the team with additional effort in developing and analyzing the tracer study as well as other aspects of the evaluation.

In Georgia, working through the ESC, **Mr. Irakli Matkava**, an expert on Georgian development, will take the lead on managing and carrying out research activities in Georgia, including interacting with MCA-G, the Georgian government, and the universities; overseeing the work of the data collection contractor; and conducting interviews and site visits.

Also in Georgia, working through ESC, **Mr. Giorgi Meladze**, an experienced manager of education institutions in the country, will conduct background research and assist Mr. Matkava with activities in Georgia.

ACT's Head of Development Projects, Ms. Sopho Chachanidze, manages the data collection activities.

Evaluation Timeline and Reporting Schedule

Table 5.1 presents an overview of the timing of major evaluation activities, reports, and dissemination.

Table 5.1. Timing of Major Evaluation Activities and Reports

Task	Task Name	Activity	Dates (due dates or activity duration)
1	Assess Evaluation Plan	Prepare Evaluability Assessment Document	Dec. 3, 2018
		Finalize Work Plan	Dec. 4, 2018
2	Develop Evaluation Design Report	Revise Evaluation Design Report to final	July 31, 2019
		Prepare Nesstar metadata template	July 31, 2019
3	Develop Evaluation Materials	Select contractor and issue contract	Feb. 15, 2019
		Obtain IRB clearance	Apr. 30, 2019
		Train enumerators	Mar.-Apr., 2019
		Pre-test instruments and procedures	Mar. 11-15, 2019
		Finalize instruments (English and Georgian)	Apr. 30, 2019
4	Prepare and Supervise Baseline Data Collection	Collect interview and focus group data	Jan. 1 – Mar. 31, 2019
		Collect survey data	Apr. 1 – Nov 30, 2019
5	Develop Baseline Report and Data Documentation Package	Survey data analysis	Aug. 15 - Jan. 15, 2020
		Qualitative analysis	Jan 1 – Sep.. 30, 2019
		Prepare draft baseline report	May 1 – Nov. 30, 2019
		Prepare Data Documentation Package	Nov. 30, 2019
		Revise Baseline Report to Final	Dec. 1, 2019 – Jan. 31, 2020
6	Disseminate	Present Baseline Report	Feb. 1 – Mar. 31, 2020

Task	Task Name	Activity	Dates (due dates or activity duration)
	Baseline Report		
7	Monitor Program Implementation and Conduct Risk Assessment	Monitor implementation	Jan. 1, 2018 – May 31, 2023
8	Revise Interim/Endline Evaluation Materials	Establish subcontract with data collection company	July 1– Aug 31, 2019
		Revise and pre-test instruments	Sep. 1 – 30, 2019, 2020, 2021, 2022
		Obtain IRB clearance	Aug. 31, 2019
9	Undertake Interim/Endline Data Collection	Train enumerators	Oct. 1 – 15, 2019, 2020, 2021, 2022
		Collect data	Nov. 1, 2019 – Apr. 15, 2023
10	Develop Final Report and Data Documentation Package	Analyze data	Nov. 1, 2022 –Jun. 15, 2023
		Prepare Draft Final Report	Jun. 30, 2023
		Prepare Data Documentation Package	Jun. 30, 2023
		Revise Baseline Report to Final	Aug. 31, 2023
11	Disseminate Final Results	Present Final Report	Sep. 1 – 20, 2023

A. Potential Programs for the Selection of the Comparison Group

University	Program	Year	Number of students accepted	Average Exam Score	Primary?
I. Javakhishvili Tbilisi State University	Applied Biosciences and Biotechnologies	2017	35	1954	
I. Javakhishvili Tbilisi State University	Applied Biosciences and Biotechnologies	2018	45	1942	
I. Javakhishvili Tbilisi State University	Biology	2015	100	1976	Backup
I. Javakhishvili Tbilisi State University	Biology	2016	100	1982	
I. Javakhishvili Tbilisi State University	Biology	2017	100	1986	
I. Javakhishvili Tbilisi State University	Biology	2018	100	1983	Backup
I. Javakhishvili Tbilisi State University	Chemistry	2015	100	1937	Backup
I. Javakhishvili Tbilisi State University	Chemistry	2016	100	1947	
I. Javakhishvili Tbilisi State University	Chemistry	2017	100	1938	
I. Javakhishvili Tbilisi State University	Chemistry	2018	100	1925	
I. Javakhishvili Tbilisi State University	Computer Science	2017	75	2052	
I. Javakhishvili Tbilisi State University	Computer Science	2018	90	2057	
I. Javakhishvili Tbilisi State University	Computer Science; Biotechnology; Geography; Geology; Electronics; Ecology	2015	295	2000	
I. Javakhishvili Tbilisi State University	Computer Science; Biotechnology; Geography; Geology; Electronics; Ecology	2016	160	2059	
I. Javakhishvili Tbilisi State University	Electronics	2018	15	1911	
I. Javakhishvili Tbilisi State University	Mathematics	2016	98	2085	
I. Javakhishvili Tbilisi State University	Mathematics	2017	61	2107	
I. Javakhishvili Tbilisi State University	Mathematics	2018	71	2003	
I. Javakhishvili Tbilisi State University	Physics	2015	79	1986	
I. Javakhishvili Tbilisi State University	Physics	2016	100	1964	
I. Javakhishvili Tbilisi State University	Physics	2017	83	1991	
I. Javakhishvili Tbilisi State University	Physics	2018	89	1981	

University	Program	Year	Number of students accepted	Average Exam Score	Primary?
I. Javakhishvili Tbilisi State University	Sciences	2018	120	1883	
Black Sea University	Informatics	2017	25	1881	
Black Sea University	Informatics (Computer Science)	2015	38	1983	
Black Sea University	Informatics (Computer Science)	2016	53	1952	
Black Sea University	Informatics (English Language Program)	2017	15	1922	
Business and Technology University	Information Technologies	2017	226	1916	
Business and Technology University	Information Technologies	2018	350	1915	
Caucasus University	Electronics and Computer Technologies	2018	33	1881	
Caucasus University	Informatics	2017	120	1923	
Caucasus University	Informatics	2018	120	1918	
Caucasus University	Informatics (Computer Science)	2015	80	1981	
Caucasus University	Informatics (Computer Science)	2016	100	1972	
Free University of Tbilisi	Computer Science and Mathematics	2015	50	2217	
Free University of Tbilisi	Computer Science and Mathematics	2016	50	2222	
Free University of Tbilisi	Computer Science and Mathematics	2017	50	2240	
Free University of Tbilisi	Computer Science and Mathematics	2018	50	2207	
Free University of Tbilisi	Electrical and Computer Engineering	2015	40	2148	
Free University of Tbilisi	Electrical and Computer Engineering	2016	40	2152	
Free University of Tbilisi	Electrical and Computer Engineering	2017	37	2177	
Free University of Tbilisi	Electrical and Computer Engineering	2018	45	2165	
Free University of Tbilisi	Physics	2015	15	2063	
Free University of Tbilisi	Physics	2016	15	2176	
Free University of Tbilisi	Physics	2017	15	2212	
Free University of Tbilisi	Physics	2018	18	2161	
Georgian Agrarian University	Biology, Chemistry, Food Technology, Agronomy, etc.	2015	180	1967	Backup
Georgian Agrarian University	Biology, Chemistry, Food Technology, Agronomy, etc.	2016	200	1981	
Georgian Agrarian University	Biology, Chemistry, Food Technology, Agronomy, etc.	2017	220	2005	

University	Program	Year	Number of students accepted	Average Exam Score	Primary?
Georgian Agrarian University	Biology, Chemistry, Food Technology, Agronomy, etc.	2018	200	2000	
Georgian Agrarian University	Electrical and Computer Engineering, Mechanical Engineering, Construction Engineering, etc.	2015	80	2022	
Georgian Agrarian University	Electrical and Computer Engineering, Mechanical Engineering, Construction Engineering, etc.	2016	80	2042	
Georgian Agrarian University	Electrical and Computer Engineering, Mechanical Engineering, Construction Engineering, etc.	2017	170	2043	
Georgian Agrarian University	Electrical and Computer Engineering, Mechanical Engineering, Construction Engineering, etc.	2018	80	2043	
Georgian American University	Construction	2016	22	1856	Backup
Georgian American University	Construction	2018	26	1839	
Georgian American University	Informatics	2016	60	1863	
Georgian American University	Informatics	2017	36	1868	
Georgian American University	Informatics	2018	49	1870	
Georgian Aviation University	Information Technologies and Flight Control Automated Systems	2017	17	1865	Backup
Georgian Technical University	Biomedical Engineering	2015	23	1863	Backup
Georgian Technical University	Biomedical Engineering	2016	25	1906	
Georgian Technical University	Biomedical Engineering	2017	40	1857	
Georgian Technical University	Biomedical Engineering	2018	40	1885	
Georgian Technical University	Chemical and Biological Engineering	2017	47	1750	
Georgian Technical University	Chemistry	2016	50	1811	
Georgian Technical University	Chemistry	2017	50	1814	
Georgian Technical University	Chemistry	2018	50	1801	
Georgian Technical University	Construction	2015	499	1861	Backup
Georgian Technical University	Construction	2016	500	1852	Backup
Georgian Technical University	Construction	2017	500	1879	
Georgian Technical University	Construction	2018	500	1848	

University	Program	Year	Number of students accepted	Average Exam Score	Primary?
Georgian Technical University	Energy and Electrical Engineering	2015	450	1858	
Georgian Technical University	Energy and Electrical Engineering	2016	450	1871	
Georgian Technical University	Energy and Electrical Engineering	2017	367	1857	
Georgian Technical University	Energy and Electrical Engineering	2018	277	1833	
Georgian Technical University	Engineering and Physics	2015	100	1828	Backup
Georgian Technical University	Engineering and Physics	2016	100	1850	
Georgian Technical University	Engineering and Physics	2017	100	1848	
Georgian Technical University	Geodesical Engineering	2017	15	1845	Backup
Georgian Technical University	Geodesical Engineering	2018	15	1901	Backup
Georgian Technical University	Geoinformatics	2018	15	1824	Backup
Georgian Technical University	Industrial Engineering	2017	30	1816	
Georgian Technical University	Industrial Engineering	2018	30	1801	
Georgian Technical University	Informatics	2015	554	1764	
Georgian Technical University	Informatics	2016	450	1799	
Georgian Technical University	Informatics	2017	423	1798	
Georgian Technical University	Informatics	2018	348	1769	
Georgian Technical University	Management Systems, Automatization and Test-Engineering	2018	30	1887	Backup
Georgian Technical University	Mathematics	2015	50	1873	
Georgian Technical University	Mathematics	2016	50	1940	
Georgian Technical University	Mathematics	2017	50	1940	
Georgian Technical University	Mathematics	2018	44	1845	
Georgian Technical University	Mechanical Engineering and Technology	2017	30	1840	
Georgian Technical University	Mechanical Engineering and Technology	2018	30	1811	
Georgian Technical University	Road Engineering	2016	30	1803	Backup
Ilia State University	Biology	2015	50	1886	Backup
Ilia State University	Biology	2016	50	1917	
Ilia State University	Biology	2017	50	1895	
Ilia State University	Biology	2018	55	1902	

University	Program	Year	Number of students accepted	Average Exam Score	Primary?
Ilia State University	Mathematics	2015	50	1919	
Ilia State University	Mathematics	2015	100	2102	
Ilia State University	Mathematics	2016	50	1915	
Ilia State University	Mathematics	2017	50	1937	
Ilia State University	Mathematics	2018	55	1888	
Ilia State University	Natural Science (Geo Sciences, Physics, Biology, Mathematics, etc.)	2015	200	1836	Backup
Ilia State University	Natural Science and Engineering faculty (Geo Sciences, Physics, Biology, Mathematics, etc.)	2016	235	1840	
Ilia State University	Natural Science and Engineering faculty (Geo Sciences, Physics, Biology, Mathematics, etc.)	2017	235	1826	
Ilia State University	Natural Science and Engineering faculty (Geo Sciences, Physics, Biology, Mathematics, etc.)	2018	185	1846	
Ilia State University	Physics	2015	15	1964	
Ilia State University	Physics	2016	15	2038	
Ilia State University	Physics	2017	15	1928	
Ilia State University	Physics	2018	15	1846	
Sokhumi State University	Biology	2016	25	1900	Backup
Sokhumi State University	Biology	2017	25	1870	
Sokhumi State University	Biology	2018	25	1895	
Sokhumi State University	Chemistry	2016	25	1804	
Sokhumi State University	Mathematics	2015	30	1879	
Sokhumi State University	Mathematics	2016	30	1906	
Sokhumi State University	Mathematics	2017	30	1857	
University of Georgia	Electrical and Computer Engineering	2016	27	1808	
University of Georgia	Electrical and Computer Engineering	2018	15	1838	
University of Georgia	Informatics	2017	29	1869	
University of Georgia	Informatics	2018	21	1851	

University	Program	Year	Number of students accepted	Average Exam Score	Primary?
University of Georgia	Informatics (Computer Science)	2016	61	1845	
University of Georgia	Informatics, Electrical and Computer Engineering, Mathematics	2015	76	1867	
	TOTAL		13,427		

Notes: All rows are in the primary sample unless labeled backup.

B. Stakeholder Comments and Responses

Note: Page numbers refer to various earlier revisions of the document.

Page Number	Comment	Response
N/A	Overall, I think that there needs to be a stronger, cohesive approach to the entire program's evaluation. At this time, it seems like there are a lot of separate pieces, but it is not clear how the questions will be answered together, the data collection efforts coordinated, and the stakeholders engaged. There are many pieces that are overlapping and this will require a significant amount of coordination and planning. This is likely realized by the evaluation team, but it is not coming through in the EDR at this point. Many of my comments below seek to get clarity on this by introducing tables that outline more clearly the specific stakeholders involved in each type of data collection and in answering each evaluation question. Additionally, I would suggest a timeline of sorts that demonstrates the proposed timing of each piece during the 5 years.	Added text at the beginning of Chapter 3 to link the questions. Also included a table that type of data being collected, sources of information and timelines in Chapter 4.
N/A	Labels/Definitions/Wording: (1) Georgia II STEM Higher Education Project is used in some places, and then without the II in other. I don't think that the II is necessary, outside of the first mention that this is the second Compact with Georgia. (2) Participants and Beneficiaries have specific definitions at MCC. Roughly, those that participate will receive perhaps training or something else from the MCC-funded program, but those that benefit will realize increases in income (a measurement for well-being). These definitions should be maintained throughout the document. (3) Related, the document uses the term 'indirect participant'. I find this to be a confusing term and given the other two definitions I think that another term will need to be used. I am not entirely clear on the concept or distinction that is intended. (4) Partnership is used on its own and typically capitalized throughout, but I am not entirely clear the intended definition. Would be good to clarify. (5) 'intakes': This term is used, but the definition is unclear - I have not seen this used in the education projects that I have worked on. Is this referring to enrollment or matriculation rates or estimates? Related, there is limited to no discussion on dropout rates. This would seem like an important data point to capture in this evaluation.	We revised the text to use the term "participant" referring to students who are the main beneficiaries. We did not use the term beneficiary in alignment with the terms being used in the template provided by MCC. We deleted the term indirect participants, and clarified what we meant by intake. We also defined partnership as between SDSU and 3 universities
pg. 10	Beneficiary Analysis: This is not fully addressed within this section. I am not sure whether it is necessary, but it is not described here. When we complete the full analysis this would consider the household as a unit and then determine the number of beneficiaries based on that, as well as considerations related to where individuals fall in the definition of poverty and their gender.	This is difficult to estimate and is not part of the evaluation. What constitutes a household for a participant in year 1 will be different in year 5 of the study.

Page Number	Comment	Response
pg. 10	Cost-Benefit Analysis: I am not the Economist who created this model, so I would need to look back at it in detail but I think that there are some key components of the model that may be missing - clarity on time horizon, major assumptions, the key parameters, etc. - and potentially some are slightly inaccurate (e.g., the end of the second paragraph in this section should end with ', ceteris paribus'). This should be a partial equilibrium models, so I am wondering how the benefit streams are actually reflected in the specific parameter calculations. We would need to ensure that there is no double-counting or projecting too far into the future on next level benefits - even if the logic model indicates that the final desired impact is to reduce poverty through economic growth. The logic model is colored differently on these aspects, but I am not clear on how to interpret that.	We added discussion of these issues.
N/A	In general, I think that there needs to be more discussion on insertion rates, labor market participation, employment/unemployment/underemployment, etc. When incomes are mentioned this is something that is within that calculation. Even if assumed I think that this should be more explicit in the EDR given its importance to the key outcomes. (particularly question 4 and 5)	We added discussion of these issues.
pg. 11	Literature Review: I am not clear on the objective of this section, is this intended to outline the plan for a literature review, be a literature review itself, or outline some of the literature that MCC/MCA completed to develop the project? This section seems to lack information on a general literature review around these potential assumptions, beneficiary streams, etc. based on the academic and practitioner's literature about similar programs. Do these seem to be well grounded within the findings in similar literature? Instead this seems to focus on the background literature for the project. That is important to review as well, but insufficient for these purposes.	Changed the title of the section to "Review of Relevant Documentation".
pg. 15	Question 4: In addition to looking to see if there is a difference in the outcomes by gender, there is more general objective of the project to increase the diversity of the student body. I am not sure if this is fully captured in the evaluation approach. Were these targeted groups actually reached during recruitment, how do applications differ by students on these types of characteristics (if it is possible to say), and then those that enroll, dropout and graduate. Perhaps this is slightly outside of the research questions or is it integrated?	The project emphasizes gender equity and targeting students with various economic backgrounds. Those are addressed in both the qualitative and quantitative pieces of the evaluation.
pg. 15	Question 5: Part of this is also to understand whether the assumptions and initial estimates that were made hold true and what learning can come from this. CBA is a way in which to determine whether this is a cost-effective program. While the ERR is the point estimate that we calculate, this is cost-benefit analysis. The point estimate will change, so focusing on the model, as a whole, is required to ensure there is learning and that the key pieces of the model are examined and discussed in this work.	We expanded the question and discussion to address the original assumptions and estimates
pg. 16	Additional details are needed on the case study to obtain a clear and complete picture of this work. Perhaps providing a bulleted list or a table of the components that will go into this work would be helpful for clarifying. The general objective of	Changes Made

Page Number	Comment	Response
	attempting to triangulate the data is clear, but how does this differ from the full report that incorporates the different sources of data collected from the various stakeholders? What will this method really provide, or is this just a reporting method?	
pg. 16 - 17	Table 3.1: (1) Data type for Question 2 seems incomplete, there will be reports and documents to review on this topic, and would it make sense to use focus groups? See comments below on overall qualitative approach. (2) Data type for Question 5: Literature review should be mentioned here, as well as the tracer study. The information collected through Questions 1 - 4 will be fed into this analysis.	Made the changes
pg. 17 - 18	Quantitative Approach - Specific comments: (1) Question 4 needs to include the skills match with employers and the employment component. (2) In a few places there are mentions of 'Year 5', but this is not clear. A timeline would be useful to understand what is happening in each year of the evaluation and how this overlaps with the programs.	Made the changes
pg. 17 - 18	Q4, Comparison Groups: (1) The score range is given, but additional information is required on the range of potential scores from this exam (overall), and whether there are general levels of performance that are considered for this exam. Without this context the range provided is not particularly useful. (2) Related, how as the range selected 1850 - 2240 for the comparison group? Clear justification is required. (3) Beneficial to explicitly state all of the entrance requirements and considerations that were made by the panel that decided on the students offered positions for these programs. This is key to developing a strong comparison group and informing the evaluation design. I did not see it mentioned anywhere. (4) The first comparison group suggest is not entirely clear. Providing some examples on the programs that you are referring to will help to explain the potential programs and individuals as compared to those in the SDSU programs. Again, perhaps it is useful to develop a table that takes that included on Pg. 8, and indicates some of the degree programs from those universities that are not covered by the Partnership, as well as some potential other universities that would be included for the comparison 2 group. (5) Based on the EDR it seems unclear on whether the method proposed will allow for a strong comparison group. There are strengths and weaknesses to this approach and it would be good to clarify this and determine if additional approaches may be required. As noted in the EDR, if they are different on observable characteristics, and these were used to determine enrollment, and were chosen because they tend to be related to unobservable characteristics that are better predictors of success then I am not sure that the statistical methods will be sufficient. Although clearly if only a test score is used, then this may also not be entirely indicative of success. More information would be helpful, and references from the literature on these considerations. (6) Did the team consider using those from previous cohorts as a comparison group - determining how to account for factors related to time, economic situation, etc.? We are doing something like this in TVET I believe. It could be an interesting avenue to explore.	We expanded and revised this discussion and provided an appendix with detailed characteristics of the comparison group.

Page Number	Comment	Response
pg. 18 - 20	<p>Survey Sample and Power Calculations: (1) State that SDSU students will be the population, but then sample of the comparison groups. However, it does not actually state the suggested sampling methodology and as the section continues it seems that the team is suggested that we basically use the population of the comparison groups as well, in order to increase the power. Please clarify. (2) Being able to measure a 25% difference in wages could be quite high. I think that there should be some analysis to clarify the level of impact that will be detected based on the assumptions within the section - given that this is a key impact. For example, what would be required to find an increase of 10%? Are we unable to get the power to measure such differences? (3) The list of assumptions provides no justification or reasoning behind those that are provided. Without additional information it is difficult to fully evaluate the strength of the proposed approach. (4) Related to comparison group comments above, the 'sample' size suggested here is quite large - 4800. I question whether this is feasible given the time and financial constraints of the evaluation. Additionally, I think that more surveying efforts may be required from what is described in the tracer study. There needs to be a more detailed discussion on weighing power with other constraints, in order to get to a reasonable 'sample' size.</p>	<p>We added more information about the assumptions and background and generally a more thorough discussion. We think it will be difficult to have power to detect below a 25 percent wage effect but given the high cost of the program it is reasonable to expect an effect of at least this size to be considered "successful". Sample size has some risks as discussed now.</p>
pg. 20 - 22	<p>Primary Data Collection: In general, this section requires more detail on the specific timing and methodology of the tracer study, as well as reference to literature to demonstrate the justification and reasoning behind the approach that is being suggested. This tracer study does not seem to follow other ones that I have seen. (1) Timing: I think that this may need to be reconsidered and grounded in the literature on when to follow-up after graduating. Sticking to year 1 and 5 seems like too few check points, the follow-up surveys noted are not clear on their objectives or how they link to the substantive topics. Perhaps sampling during certain years or the cohort approach, considering years out and a rolling basis would be useful. (2) There is only mention on surveying students and no mention of collecting related data with firms that have hired graduates. This would seem like a necessary component that needs to be tied to this survey effort (THIS SHOULD BE ADDED INTO THE QUAL SECTION). (3) Related to this, sampling and comparison group discussions: why are we not taking a cohort approach? It seems that this is considered in the analysis plan below, but not fully discussed. I would think that we would want BL/EL to be rolling depending on the cohort. They appear to be grouped together but I think that they need to be separated first. There are also economic factors that will influence these key outcomes from year to year. Grouping them all together has the potential to blur the findings some. Clearly this would decrease the power to focus on this as the primary analysis, but additional discussion could be beneficial. (4) Other related questions may be around whether they are planning to go abroad, and if so then where and why – this in the above series of questions as well. This is tied to having a better understanding of the demand for these types of graduates within Georgia and whether Georgians see this as a ticket out of the</p>	<p>We provided justification for surveying students in year 1 and year 2. We also added in this section interviewing employers to complement survey result. We included this qualitative piece here because it complements answering the outcome question. This piece does not fit in the qualitative section as the qualitative questions address implementation, partnership and sustainability. The survey does ask about whether students are going abroad, and we highlighted this issue. The survey has been piloted and we include a summary of the pilot activities implemented. Regarding comment (3), we do take into account different cohorts in the analysis. However, there isn't adequate power to look at each cohort separately.</p>

Page Number	Comment	Response
	country. (5) No mention of field testing/piloting, but I assume that this is being considered within the timelines, travel and costs.	
pg. 22 - 23	Analysis Plan: The focus here is on some of the more sophisticated components of the analysis, but doesn't provide an overall approach that ties it together. I haven't worked on PSM for a long time, so I will allow someone else to weigh in on the details, but I do have concerns with the approach due to the ability to find similar individuals within the groups that allows for the approach to work well. Given the potential weaknesses of the comparison group, I think that there should be more discussion on other types of analysis that can be done and clarity on how the team may address or get around issues with analysis when/if they arise.	We added simple regression with statistical control as an initial step of analyzing the data prior to implementing propensity score matching.
pg. 23-24	Economic Rate of Return: The ERR is the point estimate, a summary statistic of the cost-benefit analysis which has an underlying economic model built upon a series of assumption, using specific parameter - with both based on the literature, program logic, and numerous verified data sources. I provide this information to ensure that the entire model and CBA are considered within the work of estimating the ERR - the approach for data collection, literature review, etc. As such, the previous model will need to be reviewed and an outline provided of the model that will be used to answer this research question. What are the data sources? What are the key assumptions? Which assumptions from the previous model held true and which did not? Which don't we have information on yet, and how will they be used in the model used for the evaluation? What is the overall approach? Which parameters are to be included, which are not, etc.? Note time-horizon, standing, discounting rates, etc. All of this information should be mentioned within this section, as well as data sources, methods for data collection, etc.	We expanded the framing and discussion to address the original assumptions and estimates entailed in the model.
pg. 23-24	Economic Rate of Return - Costs: (1) All costs that are related to reaching the objectives of the program would be included, regardless of who incurs them. Consideration of getting costs into the same currency and ensuring that exchange rates are carefully determined. Also, ensure no double counting of costs. (2) There needs to be clarity on the specific documents that will provide the best cost estimates. There is mention of 'award details', which I am not entirely clear what this refers to, and workplan or interviews. There should be official financial statements available and used to get the best estimates of the costs. Verifying these with stakeholders will be required, but will not be the main source. I will be doing some of this exercise to complete the closeout ERRs, so we can work together with the Financial group at MCC to determine which documents to use. (3) There is a note on including opportunity costs, but the one mentioned is not clear. If the comparison is also in school then the 'without' scenario should be basically the same. If referring to something else, then good to clarify here. (4) In general, it could be beneficial to just outline in a table or bullets the key costs and benefits that will be included or indicate that this will be provided at a later date when the full review of the model is completed.	We dropped opportunity costs and added detail on the sources and types of costs.

Page Number	Comment	Response
pg. 24	Economic Rate of Return - Benefits: The approach is not entirely clear here either. Is the group in agreement with the previous benefit streams, what process will be taken to finalize the benefits that will be included, collecting the data, etc. One key question is whether some of the benefit streams are not already captured by the higher salaries and employment rates. The next level effects would seem to move away from the partial equilibrium approach – even if this is part of the program logic, it could be taking the CBA model too far or lead to double counting of benefits. As noted above, some thinking around this will be done for the closeout ERR, but input or suggestions - particularly as the literature is concerned will be helpful.	We added some more discussion of benefit streams and will consider these further in the analysis.
pg. 24 - 28	Qualitative Analysis - Overall: In general the qualitative description requires more detail on the specific methodology and approach, as well as greater clarity on the choices that are being made - why interviews, why focus groups, with whom, what type, what methods are being used for sampling, what sample size is expected, etc.?	Added a paragraph to explain the rationale for utilizing qualitative methods and when focus groups and interviews would be used. We also included a table in chapter 4 mapping each question to data source.
pg. 24-25	Qualitative - Q1: (1) There is no clear mention of trying to come up with clear metrics and methods for providing clear feedback on strength of implementation and mention where deviations from the original plans occurred and why ('within general expectation'). Even for the walk through approach. I assume that there will be a checklist or a data collection form created, but there is not much there to define the approach. (2) 'IT infra-structure': suggest deleting IT, as this seems like the incorrect terminology and confusing. (3) Could be helpful to create a table that outlines the specific components of this question that are outlined and then the method that will be used to collect data (i.e., interviews and focus groups with who?, etc.) as each one would require something different. There is not much detail here and it is difficult to get a clear picture on the various stakeholders that will be used to obtain information on each component, how they tie together, and when it will be clear that the piece of work is done. What is the role of the teachers? When have you reached saturation of information? (4) Just states that 'interview data will then be analyzed'. How will it be analyzed? How will the samples be determined? What tools, approaches, software, etc. are being used to analyze? (5) I think that there is likely to be a heavier lift in year 1, or maybe year 1-2 and then a lighter lift - at least overall, but this may differ by component. What does it really look like for the next 5 years? This may be too time intensive if done each year and quickly reduce the budget.	We clarified in the section that the Project activities are not prescriptive or well defined. So it is not feasible to figure out what the deviations were because the original plans are not specific. The purpose of the interviews was to understand how each activity was implemented and what aspects were successful and which ones were not or adjusted based on stakeholder experiences. This information is collected every year, so we can monitor change in implementation. We include more information about analysis approach under analysis section. Table 4 list data sources and number of focus groups.
pg. 25-27	Qualitative - Q2: (1) This is a good start, I appreciate the attempt to define partnership and give it some scope for how to answer the question. As noted, more work and application of the literature for this field will be required to develop a strong framework for the data collection and analysis. (2) There is reference to pg. 16 of the solicitation, but I think that there should be a full list of stakeholders provided in this document and clarity on their involvement, how many, frequency, to answer which questions, etc. This will be key in being respectful of their time and using our limited evaluation resources in the best way. (3) Again, I think that there	Information on partnership collected every year so we can monitor change over time. We include more information about analysis approach under analysis section. Table 4 list data sources and number of focus groups.

Page Number	Comment	Response
	will likely be a heavier lift in the first year and then a lighter touch. However, it should be discussed as the EDR is further detailed.	
pg. 27-28	Qualitative - Q3: (1) The answer to this question will impact the time horizon that is used for the CBA model, used to estimate the ERR. I think that these connections across the questions needs to be developed further to strengthen the overall design and use of the evaluation resources. (2) As with partnership (previous question), the word sustainable needs to be clearly defined, the components outlined, as well as the sample, data collection and analysis for each of those components laid out clearly. I think that this is a clear next/first step to designing a strong evaluation. This seems somewhat limited but there is no discussion on whether several components of sustainability were considered, and these were determined to be the focus because of x, y and z. (3) Selection of high schools: I am not clear on the proposed sampling strategy here. How will it be purposeful? Why not a representative sample? Based on what characteristics? If our goal is to increase the diversity of students then that could bias the results. Should talk to the equivalent of guidance counselors too for each of those schools selected - at least one 'adult' representative for each school. They seem to be filters for this information and we would want to know which link in the chain is weak. (4) End of the second paragraph - the last couple of sentences: This seems to be tied to the tracer study - inserted a note above about working with employers. Will need to tease out. As mentioned above, I am concerned that there are a lot of overlapping data collection efforts and there is not a clear table and process for demonstrating how the overall evaluation strategy is being brought together to reduce redundancy and increase efficiency and effectiveness of the overall approach. (5) There should be coordination with Bert on the financing for students work that he has been focused on.	Sustainability in this section is not defined in financial terms but more generally. Specifically, when examining sustainability qualitatively we are looking at awareness of STEM, visibility of the San Diego-Georgia program and the universities being successful in getting their accreditation. The section was revised to address these three points. The reason why the high schools were purposively selected is because (and with input from MCA-Georgia) not all schools encourage STEM or have STEM programs. The ones selected are those that do and have students who reflect the characteristics of students who might consider the SDSU-Georgia program or accredited STEM from the three public universities.
pg. 28	Qualitative Analysis Plan: I do not think that the plan for analysis has been sufficiently discussed below the specific questions. As noted, there should be a comprehensive qualitative plan as well as tying this together with the quantitative pieces. This is a mixed methods approach given the various methodologies that are planned to be used, but the cohesive approach that brings that together to make full use of each of them is not yet expressed here.	MCC template separates quantitative from qualitative analysis. We kept these as separate in alignment with the template. In the end the findings will be synthesized.
pg. 10; pg. 14; throughout	I'm surprised by the rather "flat" treatment of gender. While gender and increasing female participation in STEM studies is not overtly mentioned in the Program Logic per se, there is a desired output of "inclusive and Diverse student body." And as the compact developed and was implemented, increasing female participation in tertiary STEM studies became a bigger focus and goal. Several activities were undertaken to increase female student numbers (targeted recruitment, WiSci camp as a recruiting mechanism, Women's STEM club at SDSU program, etc.). While the study proposal does occasionally note that gendered differences in outcomes might be observed (i.e.. Q4, pg. 10), I would imagine that there are in fact significant	RAND built in gender disaggregation

Page Number	Comment	Response
	gendered differences at play both in terms of recruitment, retention, program implementation, but also in outcomes related to effect on wages and employment.	
	Overall, I find "students" mentioned repeatedly as a monolithic whole and I would recommend the Rand team consciously and intentionally build in gender disaggregation into all appropriate data points.	RAND built in gender disaggregation
pg. 1	MCC investment is framed as "the MCC has made a major investment to bring international higher education to Georgians by partnering three Georgian public universities with SDSU to offer bachelor's degrees in STEM disciplines." MCC and MCA don't have a direct contractual relationship with the three Georgian universities, and were not involved in SDSU's selection of the three partners. Suggest to reformulate to ensure that there is no confusion, that it was MCC who put SDSU together with the specific public universities (SDSU picked them).	Made the change
pg.2	Please clarify why HS STEM students are "indirect participants" in the project. Being aware of the project and/or part of the target recruitment group would seem to be a limited level of engagement with the SDSU program.	We deleted that
pg.2, overarching comment	Please note how international students will be integrated into the evaluation. They are not considered beneficiaries by MCC, but at some point GoG considered recruiting them a top priority. MCC also helped SDSU obtain an OFAC license to be able to enroll Iranian students, which took a considerable amount of time and effort.	We included a mention in the quantitative analysis as well as the qualitative ones.
pg.3, table	Please add citation (SDSU 45-month agreement plus date) and note which degrees were phased in later (this is noted in the source document, e.g. computer science was introduced in year 2). Please also note the relevant timeframe for degree program implementation, e.g. cohort 5 will not include all 6 degree programs.	Some information added. We don't think all this detail needs to be in this document.
pg. 3	Description of the in-person vs. distance instruction should note that the SDSU model also includes in-person U.S. faculty-led courses, with the notion that U.S. faculty presence is more significant up front, and then again to deliver certain upper-level courses. Also not sure why Georgian faculty are described as playing "a role of facilitators." They are adjunct SDSU faculty, delivering the courses, more than facilitating.	Addressed
pg. 3	What is the basis for the statement "this setback did not alter the program" with reference to lower student enrollment. Would seem to be highly questionable- I can think of a number of ways that the program was altered significantly.	Addressed
pg. 4	Please check changing verb tenses in paragraph 1.	Okay
pg. 4	Reduced education imports are not the same thing as no study abroad- please correct. We didn't set out to reduce study abroad.	Addressed
pg. 4	The ERR drew information from a variety of sources, not just the SDSU Feb. 2014 proposal. SDSU got information for said proposal from MCC and the GoG (there are many source documents)/they didn't make up the ERR in a vacuum. Would be incorrect to pin that ERR calculation on them alone.	Addressed

Page Number	Comment	Response
pg. 5 ref 5	Hope you can make good use of the wage survey documentation I sent on 2/4!	Addressed
pg. 6	Please drop "high" in "high schools" in the 3rd paragraph.	Addressed
pg. 6	I'd echo Krista's comment- we sought to enroll as many young women at SDSU as possible	Highlighted gender
pg. 6	Training for Georgian faculty was (and is) a means to the end of delivering the SDSU curricula (not a separate activity necessarily, but rather a sub-activity).	The logic model has it as its own activity, so we kept it as is
pg. 6	I would also add activities such as recruitment and popularization of STEM subjects. SDSU and MCC didn't expect to spend so much time and effort on these activities, which required a significant LOE from SDSU, MCC and MCA early on in implementation.	Addressed
pg. 7	computer certification should be changed to chemistry certification	Addressed
pg. 7, overarching comment	As mentioned when I spoke with the RAND team on 1/29, suggest integrating a line of inquiry into the dichotomy between SDSU home campus and SDSU in Georgia, looking at how this dichotomy has changed over time. This would seem to be a key element of the response to Question 2: How was the partnership established and carried out? How did it change over time?"	This is being addressed in the qualitative piece related to Q2.
pg. 9	Rita had mentioned that the Program Logic Model was drawn from the original SDSU Technical proposal- please cite here.	Cited
pg. 9	If we are adding activities, suggest to include 1) recruitment but also 2) fundraising, both activities that led to the output of "inclusive and diverse student body." Also consider adding 3) popularize STEM fields (bundled with recruitment).	Addressed
pg. 9	Don't think that "implement distance learning courses" is necessarily a distinct activity that should be over-emphasized- very few courses were taught on-line- this was a small part of delivering the SDSU curriculum.	Addressed
pg. 9	Please explain the shading out of the 3 last boxes under "long-term outcomes."	Indicates these outcomes will not be evaluated. We also mentioned that in the text.
pg. 9	When and how will adjustments noted in footnote 2 be made to the program logic?	Adjustments have now been made and the footnote deleted
pg. 9	Under "long-term outcomes" please drop the word "perusal" before "STEM graduate degrees"	Addressed
pg. 10	We are absolutely interested in the post-compact cohort of students who enroll- hope it is possible to capture this final cohort of U.S. degree program students within this evaluation.	Currently that will not be possible, but we are open to discussing an expansion of scope to capture the additional cohort.
pg. 13	SDSU's main competitor is Free University. Seems odd to limit the comparison group to the partner universities.	Comparison students also come from Free University
pg. 13	How will students who dropped out figure into the evaluation?	We will examine dropout rate as a descriptive indicator. However, it will not be included in the main analysis
pg. 13	Average monthly wages of \$1000 in treatment seems very low. How was this as well as the average comparison group wage of \$800 determined?	Comparison group wages taken from the wage study which predicted a 44 percent increase for SDSU

Page Number	Comment	Response
		graduates. We used a more conservative 25 percent to power the study.
pg. 20	Please review Dalberg studies and analysis about other branch campuses and similar models in international higher education (from 2015, 2016).	We have not located these articles. Can you provide them?
pg. 21	In order to "unpack" question 3, it's important to have the full context, e.g. a transition out of U.S. degree delivery was always anticipated, therefore the goal was never to sustain U.S. degree programs ad infinitum. Suggest we all spend some time figuring out exactly what we mean here, e.g. starting by asking, what activities were intended to be sustainable, and what were intended to be temporary/short-term during the life of the Compact? It's also worth recognizing that from day 1, we didn't expect the SDSU programs to be sustainable until compact year 7 (another reason to include cohort 5 in the analysis).	Addressed
pg. 21	Should be sure to interview the Center for International Education, as SDSU sub-contracted Georgian HS recruitment to them. Also obtain list of SDSU "Feeder Schools," or high schools that signed MoUs with SDSU.	Addressed
pg. 21	What is the "thematic analysis" to analyze the data collected through interviews and focus groups?	Elaborated on the qualitative analysis section
pg. 21	Regarding financial viability, please make time to review the SDSU budget in person with SDSU's financial director Lado. As the RAND team said, it will be very challenging to understand a giant spreadsheet- really need SDSU budget person to explain.	We have had an initial discussion and will have further conversations.
pg. 21	Who makes up the "Education Strategy Center" team?	clarified in Chapter 5.
overarching comment	One major element of this program is delivering the "American experience" in education, e.g. plagiarism is not allowed, there is an early application process, students are expected to attend and participate in class, students are supported in pursuit of internships and careers, faculty have office hours and are available to students after hours etc. How can this very key aspect of the SDSU program in Georgia be integrated into the evaluation?	Integrated into survey questions and qualitative piece.
overarching comment	As alluded to above, we (MCC, MCA, SDSU) spent a significant amount of time and effort doing student recruitment and fundraising. How can this aspect of the SDSU program in Georgia be integrated into the evaluation?	Will be addressed as part of Q1
overarching comment	Perhaps I missed it, but are there also plans to interview SDSU faculty trained who aren't teaching, who are teaching, etc.? Seems most of the interviewees are students and businesses.	We are interviewing them
overarching comment	We will now have at least one "gap year" between SDSU U.S. degree program enrollment (last group fall 2019) and Georgian partner university ABET-accredited degree program enrollment (earliest fall 2021). Please consider how we integrate this relatively significant development into the evaluation.	We will have to address this over time as the situation becomes more clear.
overarching comment	Not to be nit-picky but I noticed somewhat interchangeable usage of the following words: initiative, partnership(s), project, program.	Aligned the terms

Page Number	Comment	Response
overarching comment	This looks to me like a thorough plan for assessment, and I have no reservations about the proposed process or metrics. My guess is that the issue of sustainability is going to be the hardest to establish, but I think we know that already.	We agree.
overarching comment	<p>I find these foci realistic in that the available data, time, and resources could speak to them. Although the ultimate (distal) goals have been often discussed – effect on the Georgian economy – the evaluation plan authors rightly recognize that these effects could only be speculated about.</p> <p>I find the evaluation framework, methodology, and analytical plan sound.</p> <p>If I would like to clarify anything, it would be a goal of assisting Georgian partners in becoming ready for (ABET, ACS) accreditation and not in obtaining accreditation. The latter is beyond our control.</p>	We clarified this point in the report
overarching comment	This looks to me like a sound plan. Question 3 of sustainability, surprisingly, does not include any financial outlook. If included, though, it should be a responsibility of the Georgian partner universities, governmental agencies, and private donors. The logistics of sustainability may include also the program's "vertical" expansion towards collaboration in the sphere of graduate studies (the plan mentions some elements of this in other sections.)	In the sustainability section we included the examination of indicators of financial sustainability to provide general information on whether the universities are likely to be able to sustain the program.
overarching comment	<p>Question 3 - Sustainability</p> <p>To me the financial sustainability has always been and remains to be a big question. I am not sure if RAND plans to evaluate that.</p>	In the sustainability section we included the examination of indicators of financial sustainability to provide general information on whether the universities are likely to be able to sustain the program.
overarching comment	<p>Question 5 - post compact economic rate of investments (ERR)</p> <p>As the report pointed out, the original proposal estimated to have over 2000 students by year 6 but the actual number of students is more around 600. There are many factors that caused the discrepancy such as family's ability to pay and the number of Georgian high school graduates who are interested in STEM and meet the language requirement. As a result, the model has changed from SDSU recruiting 500 new students every year for many more years post compact, to having only five SDSU cohorts and accelerating the timeline of partner universities getting ABET and ACS accreditation.</p> <p>If I understand correctly, RAND Corp. plans to look at only the 5 cohorts of SDSU graduates when calculating the economic rate of return of the \$30 million investment. However, there are economic benefits from the post-compact cohorts who will enroll into the partner universities' soon-to-be-accredited STEM programs. These students will be taught by SDSU trained Georgian faculty and they will study in SDSU built labs and building. The training provided to the Georgian faculty and renovation of labs and construction of the ISU building were funded by the \$30</p>	We agree, we will use information on post compact cohorts established during the RAND study.

Page Number	Comment	Response
	million grant. So I would recommend that RAND Corp. expands their scope of evaluation when calculating ERR to include these future cohorts.	
overarching comment	(1) we will "assist" our partner universities to build the components needed to apply for accreditation/certification, but they have to actually implement, maintain and sustain independently, and (2) they will need financial support from the Georgian Government to maintain all the new infrastructure and expensive instruments we installed in order to secure and maintain accreditation/certification (along with a few other suggestions).	We addressed these issues in the revised version.
Table 1	Replace in table under TSU "Chemistry (Biochemistry focus)" with "Chemistry" "Biochemistry" in 2 lines since they are different just like EE and CE Under GTU Remove "Chemistry"	Made the changes
pg. 5	I note a very careful wording (i.e., "were presumed to absorb...."). Has anything been done to test/support the assumption? (e.g., observation and/or evaluation done by the SDSU mentor faculty, anything else?).	Changed the wording and focused it on behavior
pg. 8	It is hard to comment this without the original analyses; however, based on what we know about the resources needed to, among others, maintain the same instruction and equipment, the estimate of \$1,589 is likely too low.	So noted.
pg. 8	Establishing a reference/comparison point to test this hypothesized benefit is critical. I presume more will be said about this later in the text.	See methodology section
pg. 11	All this stands to reason. Any interpretation must take into account initial demand- and paying-capacity studies, which, as stated before, had some serious flaws.	Agree
pg. 12	Also, perhaps with different partner universities.	Added suggestion
pg. 13	This is critical! Whereas I understand MCC's interest in post-compact developments– indeed, it speaks to the critical question of a one-time shot versus a lasting effect – RAND is right to point out that it would require a whole new evaluation.	N/A
pg. 13	Whom will they interview given a huge turn over in the Ministry? Caution is necessary not to take the most recent/available respondents as qualified (have sufficient information/participation) to comment on the entire process.	Made revisions to address the turnover.
pg. 15	How operationalized?	We cover in detail in the methodology section
pg. 18	Not sure what this implies for the SDSU-G cohort that has just graduated.	They are included. Clarified wording.
pg. 23	One could additionally speculate about higher tax revenue from increased earnings (assuming progressive taxation in G.).	Added.

Page Number	Comment	Response
pg. 24	So, it is important that all involved parties recognize this and accept that a primary focus will be on individual-level outcomes.	We agree.
pg. 24	Has anyone on the SDSU-G team participated in the creation of these check-lists?	The checklist was based on the activities defined by SDSU's logic model. It was a simple list of what should be in place.
pg. 25	My understanding is that this was a collaborative process that involved MCC, and MCA. Correct? If so, it should be recognized in the implementation evaluation plan.	We made revisions to the text here. And yes when looking at partnerships we will look at the role of MCC and MCA as well.
pg. 31	When will SDSU have access to data? There will be a wealth of information in these data that our researchers, in collaboration with their Georgian counterparts can "mine" and expand on.	The data will be provided after the RAND study is completed.

References

- Constraints Analysis, Georgia, Final Draft.* Government of Georgia, July 2011.
- Demand Study Relating to the Development of Millennium University, Revised Report.* Texas International Education Consortium, May 2011.
- Hans Gutbrod, CRRC, “Measuring Wealth Differential.” Presentation for MCC, 8 November 2011.
- Karam, R, Pane, J., Griffin, B.A., Examining the Implementation of Technology-Based Blended Algebra I Curriculum at Scale Cognitive Tutor Algebra 1 Program in Middle and High Schools. *Educational Technology Research Development*, 2016, 65(2), 399-425.
- Monitoring and Evaluation Plan, Compact II. Millennium Challenge Account – Georgia, Version 3*, January 2018.
- Solicitation: Independent Evaluation Services in support of the Georgia II STEM Higher Education Project*, Millennium Challenge Corporation, May 15, 2018.
- Technical Proposal for Selection of Partner Institution(s) for Capacity Building and Establishing Bachelor Degree Programs in Science, Technology, Engineering and Math Higher Education in Georgia: A Partnership between San Diego State University, Ilia State University, and Tbilisi State University.* San Diego State University, March 2013.
- Vernez, G, Karam, R., Marshall, J, School-Based Management in Indonesia, Santa Monica, CA: RAND, 2012.